UNIVERSITY OF ALBERTA

CHARACTERIZATION OF THE ENTEROBACTERIAL *frd-efp* GENE INTERVAL: IMPLICATIONS FOR THE ORIGIN OF ANTIBIOTIC RESISTANCE AND VIRULENCE

BY

RUSSELL EDGAR BISHOP

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

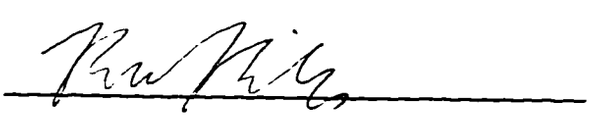
EDMONTON, ALBERTA

FALL, 1997

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR:	Russell Edgar Bishop
TITLE OF THESIS:	Characterization of the Enterobacterial *frd-efp* Gene Interval: Implications for the Origin of Antibiotic Resistance and Virulence
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED:	Fall 1997



14556 107A avenue
Edmonton, Alberta
CANADA
T5N 1G5

Date: 09/25/97

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled CHARACTERIZATION OF THE ENTEROBACTERIAL *frd-efp* GENE INTERVAL: IMPLICATIONS FOR THE ORIGIN OF ANTIBIOTIC RESISTANCE AND VIRULENCE submitted by RUSSELL EDGAR BISHOP in partial fulfillment of the requirement for the degree of DOCTOR OF PHILOSOPHY.

Dr. Joel H. Weiner [Supervisor]

Dr. Michael N. G. James

Dr. J. N. Mark Glover

Dr. Susan E. Jensen

Dr. Milton H. Saier Jr.

Date: 2/1/97

Concordia
Domi
Foris
Pax

For Kate,

the Pea,

and our families.

ABSTRACT

The enterobacterial *frd-efp* chromosomal gene interval comprises seven distinct open reading frames in *Citrobacter freundii*, from which two regulatory genes are deleted in *Escherichia coli*. The genes encode the AmpC class C ß-lactamase (and the neighboring AmpR regulatory protein in *C. freundii*), an outer membrane lipoprotein that belongs to the lipocalin protein family (Blc), a member of the small multidrug resistance protein family (SugE), the entericidin lipoproteins (EcnA and EcnB) and the neighboring response regulator (EcnR) of *C. freundii*. On the basis of a structural similarity between penicillin and an LD-peptide bridge in the murein, it was proposed that the AmpC protein may function as an LD-endopeptidase in addition to its known function as a ß-lactamase. Despite the identification of a growth defect in an *ampC* deletion mutant, murein analysis disproved the LD-peptidase hypothesis. However, a corollary that AmpC induction occurs indirectly via a recyclable cell-wall peptide is now established. This depended on an AmpR purification procedure, which demonstrated that AmpR interacts with its operator as a pair of dimers. The *blc* gene was identified immediately downstream of *ampC* and shown to be expressed in stationary phase under the control of *rpoS*. Overexpression, membrane fractionation, and metabolic labeling with [^{3}H]-palmitate established that Blc is an outer membrane lipoprotein. The amino acid sequence of Blc is homologous with an ancient group of lipocalins that articulate with membranes through hydrophobic anchors. Similarly, it was demonstrated that subunits II of the cytochrome oxidases were adapted to articulate with membranes by lipoprotein anchors. A stationary phase bacteriolysis phenotype was traced to a novel bacteriolytic module, called the entericidin locus, which is activated by *rpoS* and repressed by the osmoregulatory *ompRenvZ* signal transduction pathway. The entericidin locus encodes tandem paralogous genes, which are conserved among the enterobacteria, and which direct the synthesis of small cell envelope lipoproteins. Both entericidins adopt amphipathic α-

helical conformations and modulate membrane properties. The first gene product, entericidin A, is an antidote to the second bacteriolytic gene product, entericidin B, which is consistent with an emerging paradigm of programmed bacterial cell death.

ACKNOWLEDGEMENTS

I must express my gratitude to my supervisor, Joel Weiner, for supporting all of my scientific endeavors, even those that turned out to be wrong. Joel once said to me, "you must will your experiments to work." Perhaps I have now come to understand what that really means. At times, my endeavors took me away from Joel's lab, but he always continued to support me from afar. My first journey was to the laboratories of Stewart Cole at the Pasteur Institute, and Jean van Heijenoort at Orsay in Paris. One year in Germany was spent in collaboration with Joachim-Volker Höltje at the Max Planck Institute for Developmental Biology in Tübingen. One month in Vancouver was supported by Robert Hancock of the University of British Columbia. At the University of Alberta, I performed several experiments in the laboratories of Brenda Leskiw, Luc Berthiaume, Cyril Kay, and Robert Hodges.

Members of my committee, and other investigators, were instrumental in facilitating the development of my ideas. Michael James and Randy Read brought out the third dimension that I was usually seeking. I thank Bernie Lemire for providing criticism without cynicism. Susan Jensen reviewed the final chapter before it was submitted for publication and Mark Glover kindly took the place of Bernie Lemire while he was on sabbatical. I also thank Laura Frost and Brenda Leskiw for introducing me to the RNA world, and to Dennis and Jean Vance for introducing me to the lipid world. The underlying theme, however, was membranes, and here I benefited tremendously from the members of Joel's laboratory and the MRC Group in the Molecular Biology of Membranes. In particular, I must thank Raymond Turner, Richard Rothery, Larry Fliegel, Carol Cass, and Marek Michalak for their insights. Technical support from Gillian Shaw, Amedchand Boodhoo, and Susanne Pawliuk was instrumental, both literally and figuratively. The numerous investigators who provided bacterial strains and plasmids cannot be listed here, but are mentioned throughout the main chapters. Thanks.

I was supported by a Faculty of Medicine 75th Anniversary Studentship, a Graduate Studentship from the Alberta Heritage Foundation for Medical Research, an Annual Grant from the German Academic Exchange Service, and a Dissertation Fellowship from the Faculty of Graduate Studies and Research.

Finally, I must thank my wife, Kathryn Carter, for teaching me how to properly use a semicolon; she is my greatest discovery and, as we enter into parenthood, she shares in my greatest experiment. I wish to dedicate this thesis to my parents, without whose support none of this would have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

A_2pm	diaminopimelate
ADP	adenosine-5'-diphosphate
ApoD	apolipoprotein D
ATP	adenosine-5'-triphosphate
Ap	ampicillin
CD	circular dichroism
Cm	chloramphenicol
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FPLC	fast protein liquid chromatography
Gdn/HCl	guanidine hydrochloride
GlcNAc	N-acetylglucosamine
HPLC	high performance liquid chromatography
HTH	helix-turn-helix
HDL	high density lipoprotein
IPTG	isopropyl-ß-D-thiogalactopyranoside
kbp	kilobasepairs
kDa	kilodaltons
Kn	kanamycin
LB	Luria-Bertani medium
LSB	Laemmli solubilization buffer
MOPS	3-[N-morpholino]propanesulfonic acid
MurNAc	N-acetylmuramic acid
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis

PBS	phosphate buffered saline
PCR	polymerase chain reaction
pmf	proton motive force
rbs	ribosome binding sequence
SDS	sodium dodecyl sulfate
Taq	DNA polymerase from *Thermus aquaticus*
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCA	trichloroacetic acid
TEMED	tetramethylethylenediamine
Tet	tetracycline
TFE	trifluoroethanol

CHAPTER 1

INTRODUCTION

The widespread use of antibiotics for the treatment of bacterial diseases has created enormous pressure for the selection of antibiotic resistant organisms (1). The evolution of antibiotic resistance determinants was initially deemed unlikely due to low mutation frequencies in bacteria. However, it is now clear that a pre-existing pool of antibiotic resistance genes in the natural environment can be readily acquired and transferred between bacterial species with extraordinary facility. Currently, antibiotic resistance has become a crisis in hospitals where many formerly indispensable antibiotics are now useless. This has prompted a renewed effort to understand the nature of antibiotic resistance with the hope of developing new approaches for antibiotic discovery and for the prudent use of antibiotics.

The failure to anticipate the rise in antibiotic resistance reflects our failure to recognize the power of evolution in microbial populations. Until recently, microbiology has been one of the least evolutionarily oriented of the biological disciplines, perhaps due to the mistaken notion that extensive gene exchange between species implies that bacteria share a single gene pool (2). Present views indicate that bacterial species are largely independent with respect to their chromosomal genes, but a variety of insertion sequences, transposons, plasmids, and phage, which were once regarded as adaptive mechanisms for recombinational gene exchange, are better regarded as parasitic byproducts of recombination (3). This necessitates the partitioning of the bacterial genome into euchromosomal and accessory, or parasitic, genetic elements.

1997 marks the completion of the *E. coli* genome sequencing project, and the beginning of a new era in the cellular and molecular biology of the world's most extensively characterized organism. The first edition of the book 'Escherichia coli and Salmonella typhimurium: cellular and molecular biology', which was published in 1987

(4), coincided with my first summer studentship of undergraduate research on *E. coli*. The second edition of that book was published in 1996 (5) and concludes with an epilog which notes that only about half of the genes of *E. coli* have been identified with a physiological function. "The big question is what is hiding in that half of the *E. coli* genome? What are all these genes?"

Traditionally, the characterization of *E. coli* has involved the identification of a biochemical activity followed by a search for the corresponding gene. The analysis of function has typically preceded any ancillary consideration of a gene's evolutionary origin. In the post-genomic age, one is invariably struck by the extent to which evolutionary relationships can be deduced in the absence of any knowledge of gene function. These evolutionary relationships are often the initial basis from which functional investigations now proceed. Dobzhansky once stated (6) that "Nothing in biology makes sense except in the light of evolution." Perhaps it can also be said that nothing in evolution makes sense except in the light of molecular biology. In this regard, I agree with the statement of Richard Dawkins (7, p. 162) that "molecular biologists might be aided in their research by the realization that DNA is not working for the good of the cell but for the good of itself." This paradigm of "gene-selectionism" will be treated in greater detail in the Discussion (Chapter 9).

I have investigated the structure, function, and regulation of a cluster of genes located in the *frd-efp* interval near 94.5' on the *Escherichia coli* chromosome and compared them with the closely related *Citrobacter freundii*. I will demonstrate that each of the genes in the cluster encode components of the bacterial cell envelope, and that several contribute, either directly or indirectly, to the antibiotic resistance gene pool. An outline of bacterial evolution precedes an introduction to the components in the *frd-efp* interval.

The tree of life

Although Leeuwenhoek discovered microorganisms in 1675, it was Haeckel in 1866 who first established a separate kingdom of microorganisms, called protists, which were distinguished from plants and animals (8). A protist grouping called monera included all known bacteria except the blue green algae, which were mistakenly believed to be primitive plants. The modern "Five Kingdoms" of life included animals, plants, fungi, protists, and monera (9). The monera were separated to distinguish organisms that contain a nuclear membrane (eukaryotes) from predecessors that lack such a membrane (prokaryotes). Most of the diversity of life on earth was believed to be represented by the eukaryotes, and the prokaryotes were regarded as a monophyletic group that simply lacked several fundamental cytological properties of eukaryotes (10).

As noted by Woese (11), early efforts to classify bacteria according to their morphological features, which had been so successful with plants and animals, were discredited by distorted phylogenies that led many to "deemphasize the role of evolutionary considerations in the development of microbiology." However, as early as 1962, J.D. Bernal (12) noted "that there is an underlying *unity* in biochemistry which implies a *biochemical evolution* that gave us all the various forms, performances, and behaviors of the plants and animals of today." A 1965 article entitled "Molecules as Documents of Evolutionary History," by Zuckerkandl and Pauling (13), influenced many evolutionary biologists, and microbiology would soon be transformed by the evolutionary implications of molecular sequence data. The formulation reached by Carl Woese, on the basis of small subunit ribosomal RNA sequence comparisons, includes three lines of evolutionary descent: Eukarya, Archaea, and Bacteria (14). This scheme has been vindicated by the completely sequenced genomes of members from each of the three domains (15,16). It is now clear that the eukaryotes provide only a small part of the biological diversity on the planet and that the prokaryote-eukaryote dichotomy is

intrinsically flawed: the Archaea contain features of both the Bacteria and the Eukarya (17).

The fact of evolution is that all life on earth descended from a common ancestor, or cenancestor. The Universal Tree of Life is based on the comparison of homologous macromolecules, i.e. those that share a common ancestor (18). Orthologous macromolecules share common functions among different organisms. However, paralogous macromolecules arise from gene duplication events and are adapted for different functions (19). The Universal Tree, which is based on orthologous structures, can be rooted by comparing trees derived from paralogous structures that were initially duplicated in the cenancestor (20,21,22). The universal tree has been rooted on the bacterial line, which revealed that the Archaea and Eukarya are sister groups with a common history exclusive of the Bacteria. However, the endosymbiotic origin of mitochondria and chloroplasts proposed by Margulis, and now established, indicates that these eukaryal organelles are derived from Bacteria (23,24,25). The mitochondria are derived from the α-subdivision of proteobacteria and chloroplasts are derived from cyanobacteria (11). The Eukarya must have acquired these organelles by a process of horizontal transmission (23). Therefore, evolutionary relationships based on homologous macromolecules can break down if one fails to account for horizontal and vertical transmission or to distinguish orthologous and paralogous structures.

A consequence of having a rooted Universal Tree of Life is that those organisms clustered near the root of the tree are more closely related to the cenancestor. Most of these are Archaea and Bacteria known as thermophiles because they grow at temperatures above 80C (16). It seems that life on earth may have originated under conditions similar to those occupied by present-day thermophiles, which are often isolated from hot springs and hydrothermal vents on the ocean floor (26). In any case, it is no longer valid to presume that *Escherichia coli*, the best characterized organism in biology, is somehow representative of prokaryotes in general. In the Universal Tree, *E. coli* is placed among

the γ-subdivision of proteobacteria, also known as purple bacteria, which are closely related to mitochondria (11). We must carefully consider whether homologous structures shared between *E. coli* and the Eukarya resulted from horizontal or vertical transfer.

This dissertation is concerned with a branch of Gram-negative organisms known as the *Enterobacteriaceae*. The model organism is *Escherichia coli* and it will be compared with its close relative *Citrobacter freundii*. At about 45% DNA relatedness, these organisms are as closely related to each other as they are to *Salmonella* (27). The predicted time of divergence of *Escherichia* and *Salmonella* is estimated at 120-160 million years ago, in correspondence with the origin of mammals. The *Enterobacteriaceae* are Gram-negative facultative anaerobes that ferment the milk sugar lactose, which was a mammalian invention. While *Salmonella* is a pathogen that typically parasitizes mammals, *Escherichia* is an opportunistic pathogen that normally exist as a harmless commensal member of the intestinal flora of mammals (28). *Escherichia* is a major component of the large intestine flora, but is a minor component of the total intestinal flora. The estimated doubling time in the intestine for *Escherichia* is one day, where laboratory cultures double every 20-30 minutes in rich medium. Secondary habitats of *Escherichia* are soil and water where it is present as fecal contamination and can survive for a few days without dividing (27). *Citrobacter* is better adapted to occupy extraintestinal environments since it can be isolated from soil in the absence of fecal contamination (29). Therefore, it appears that these three genera of *Enterobacteriaceae* are differentially adapted to occupy two distinct habitat types and this may be reflected in physiological differences.

The Enterobacterial *frd-efp* gene interval

The fumarate reductase operon (*frd*) of *E. coli* encodes four protein subunits of the fumarate reductase complex. The *frd* operon is located at 94.5 minutes on the *E. coli* chromosome and is immediately followed by the *ampC* gene (Figure 1-1), which encodes

the AmpC ß-lactamase (30). The promoter of *ampC* transcription resides within the *frdD* cistron and must read through the *frd* Rho-independent terminator, which acts as a growth rate-dependent attenuator of *ampC* transcription known as *ampA* (31). In *C. freundii*, the *frd* and *ampC* operons are separated by an additional divergently oriented gene known as *ampR* (32), and the Rho-independent terminator between *frd* and *ampR* is bi-directional. The AmpR protein belongs to the LysR family of transcriptional regulators and renders the neighboring *ampC* gene inducible by ß-lactam antibiotics (33). The induction process depends on several unlinked genes, which include *ampD* and *ampG*. Mutants defective in *ampD* constitutively overproduce AmpC (34), while *ampG* mutants are constitutively repressed (35).

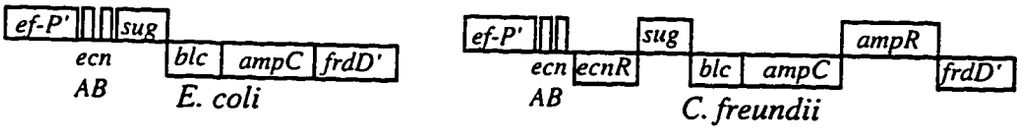


Figure 1-1. The enterobacterial *frd-efp* gene interval.

Immediately downstream of *ampC*, in both *E. coli* and *C. freundii*, lies an open reading frame that encodes a member of the lipocalin family of proteins, which have only previously been identified in Eukarya (36). The Bacterial lipocalin (Blc) is encoded by the *blc* gene, which converges on another open reading frame with no Rho-independent terminator between them and with the translational termination codons overlapping. The convergent gene, known as *sugE*, encodes a member of the small multidrug resistance (Smr) family of proton motive efflux pumps, but the specific function of SugE is unknown (37). Located adjacent to *sugE* in *C. freundii* is another divergently oriented regulatory gene (*ecnR*) that is absent from *E. coli*, which in this case belongs to the OmpR family of response regulators (38). The Rho-independent terminator of the response regulator is bi-directional with another small open reading frame that encodes a lipoprotein. The small lipoprotein gene is preceded by a paralog (*ecnA* and *ecnB*), resulting from an apparent gene duplication event in both *E. coli* and *C. freundii*; these

are located immediately downstream of the Rho-independent terminator for the upstream operon (*efp*) that encodes translational elongation factor P (39).

Altogether, seven open reading frames in *C. freundii* and five open reading frames in *E. coli* are specified by the *frd-efp* gene interval. Homologous *Eco*RI fragments bearing the *frd-efp* interval are 7.6 kbp in *C. freundii* and 6.1 kbp in *E. coli* (40). I investigated the function and regulation of the AmpC ß-lactamase by comparing and contrasting the *E. coli* and *C. freundii ampC* operons. This led to the unexpected discovery of a stationary phase bacteriolysis phenotype that redirected efforts toward the genes downstream of *ampC*. As a prelude to these investigations, the structure-function relationships of the protein families relevant to the individual members of the *frd-efp* interval are discussed.

LysR and OmpR families of transcriptional activators

Transcription is governed by RNA polymerase. In bacteria, RNA polymerase recognizes promoter elements in DNA by virtue of the σ-subunit (41). The primary σ-subunit expressed by growing *E. coli* is encoded by the *rpoD* gene and encodes a 70kDa protein (σ^D). A consensus DNA sequence for σ^D recognition is centered around hexameric sequences at the -10 and -35 positions with respect to the transcription start site of a given promoter: TTGACA-N_{17}-TATAAT (42). This is a hybrid promoter of optimal strength that was created by fusing the -35 region of the *trp*-promoter with the -10 region of the *lacUV5*-promoter and is known as a *tac*-promoter. Compared with the *tac*-promoter, most σ^D-promoters *in vivo* are degenerate in accordance with lesser promoter strength.

Under starvation conditions, *E. coli* produces an alternate σ-factor encoded by the *rpoS* gene (43), which encodes a 38kDa protein (σ^S). The set of genes required under starvation conditions that are recognized by σ^S have a distinct consensus heptad sequence centered around the -10 region with respect to the start site of transcription:

CTATACT (44). While no -35 region is necessary for σ^S-dependent promoters, a region of intrinsic DNA curvature specified by phased adenine-tracts are usually observed upstream. Some σ^S-dependent promoters are recognized by σ^D under growth conditions because the σ^S consensus can be compatible with that of σ^D (45). Several cellular signals for control of *rpoS* expression have been identified. These include guanosine-3',5'-bispyrophosphate (ppGpp) of the stringent response, which is produced in response to amino acid deprivation, and a homoserine lactone, which is a quorum-sensing metabolite produced in response to high cell density.

Alternate sigma factors provide a facile means to modulate global gene expression in response to environmental conditions. σ^D and σ^S represent the two main σ-subunits under vegetative conditions, but minor σ-subunits are also evoked in response to specific environmental cues. However, the cell more frequently utilizes ancillary transcription factors to govern transcription in response to specific signals. The most prominent transcription factors in Bacteria are the OmpR-family of response regulators (38) and the LysR-family of transcriptional regulators (46). Generally, both families of transcription factors control RNA polymerase by activating transcription in response to environmental signals, and bind to DNA by virtue of the helix-turn-helix motif.

The helix-turn-helix motif was first identified in the λ-cro protein and the catabolite gene activator protein (CAP), also known as CRP (cyclicAMP-dependent regulatory protein.) Briefly, the regulator interacts with DNA as a dimer and projects an α-helix (the reading helix) into adjacent major grooves of B-form DNA (47). In the LysR family, the specific consensus of the cognate DNA operator sequences are manifest as inverted repeats, which predicts that the LysR dimers are placed in a back-to-back orientation (46). In the OmpR family, the consensus sequences manifest themselves as direct repeats, which predicts a back-to-front orientation of the dimer (38). Furthermore, the helix-turn-helix motif is found in the N-terminal domain of the LysR family and in the C-terminal domain of the OmpR family.

The LysR family typically responds to environmental signals in the form of specific low molecular mass effectors. Since each LysR member binds a distinct ligand, the ligand binding domain in the C-terminal regions have distinct structural features. LysR proteins are transcribed divergently from the structural genes under their control, and bind to operator sequences that overlap with their own promoters (46). Therefore, most LysR proteins provide autoregulation since they bind their operators in the absence of activating ligand. In contrast, the OmpR family uses a common regulatory mechanism of covalent modification by phosphorylation of a conserved aspartic acid residue. The degree of operator interaction is proportional to the degree of phosphorylation, which is a graded effect of kinase and phosphatase activity (48). The labile acyl-phosphate is produced by a cognate histidine-protein kinase, which serves as the environmental sensor. The basic requirement for a sensor-kinase and a response-regulator is sometimes referred to as a two-component regulatory system. However, the labile phosphoramidate in the histidine-protein kinase can be transferred to multiple response regulators to comprise complex regulatory networks.

The murein hydrolases

D-alanine is a component of the peptidoglycan layer known as the murein sacculus, which forms a bag-like exoskeleton that determines the structure of the bacterial cell and stabilizes the cytoplasmic membrane against high internal osmotic pressure. The enzymology of D-alanine metabolism is of tremendous importance in antimicrobial chemotherapy. In the final steps of murein biosynthesis, glycan chains of peptidoglycan are polymerized and crosslinked together by bifunctional transglycosylase-transpeptidase enzymes (49). The transpeptidase domains can be inactivated by penicillin, which is a structural analog of the C-terminal acyl-D-alanyl-D-alanine component of peptidoglycan, and are known as penicillin binding proteins (PBPs). In addition to transglycosylation and transpeptidation, PBPs can function as D-alanyl

carboxypeptidases or D-alanyl endopeptidases. Murein hydrolase activity is generally regarded as a mechanism for the modulation of murein synthesis during cell growth and division (50).

ß-lactamases hydrolyze the ß-lactam nucleus of penicillins and cephalosporins (Figure 1-2). The ß-lactamases are grouped into classes A, B, C, and D according to amino acid sequence similarity. Class B ß-lactamases are metalloenzymes and classes A, C, and D belong to a homologous family which include D-alanyl peptidases and a D-aminopeptidase (51). All members of this larger superfamily are serine amidases, which react with substrate to esterify an active site serine residue thus forming an acyl-enzyme intermediate that can be transferred either to water (hydrolysis) or to an amino group (transpeptidation). An absolutely conserved residue surrounding the active site serine is a lysine positioned three residues downstream. In the determined crystal structures of class A and C ß-lactamases (52,53), the active site serine lies at the base of a helix with the ε-amino group of the lysine positioned on the same face of the helix within hydrogen-bonding distance of the serine OŸ. It has been proposed that the pKa of the lysine (which is buried in the hydrophobic core of the enzyme) is sufficiently low that it exists in its deionized state under physiological pH and functions as a general base by removing a proton from the serine hydroxyl-group (53). This is activated for nucleophilic attack on the carbonyl carbon of the ß-lactam (and presumably acyl-D-alanyl-D-alanine in the case of PBPs) to form a tetrahedral intermediate, which is stabilized by hydrogen bonding interactions with the backbone amides of residues that surround an oxyanion hole. The proton abstracted by the lysine is delivered to the leaving group nitrogen atom of the substrate via a suitably positioned hydroxyl group, which can belong to either a serine or a tyrosine (53,54).

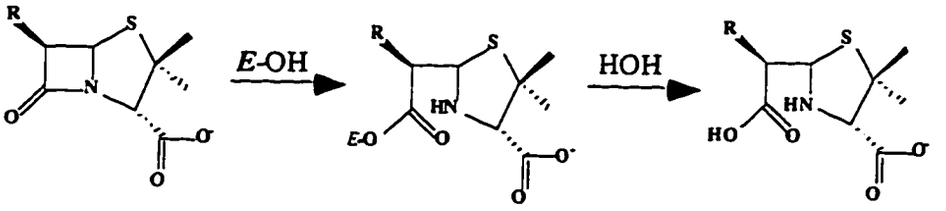


Figure 1-2. Hydrolysis of ß-lactam antibiotics by serine ß-lactamases (53).

Once the acyl-enzyme intermediate has been formed in the class C ß-lactamases, a hydrolytic water molecule can move into the position formerly occupied by Oγ of the catalytic serine and become activated for deacylation by the lysine (54). However, the class A ß-lactamases possess an adaptation known as the Ω-loop, which activates the hydrolytic water molecule with a glutamate residue that serves as a distinct general base for deacylation. An earlier postulated role for the glutamate associated water molecule in acylation fails to account for the importance of the lysine or the absence of the Ω-loop glutamate in class C ß-lactamases (55). These authors support their conclusions by demonstrating that a titratable lysine with lowered pKa could not be detected by NMR in a class A ß-lactamase (56). However, the pKa of the lysine may still be lowered in the Michaelis complex owing to interactions with the ß-lactam substrate. More recently, a number of distinct enzymes have been identified in which a lysine is implicated as a general base (57), which supports the idea that a lysine could play a similar role in ß-lactamases.

The lytic transglycosylases (50) are a group of exomuramidases that cleave the ß-1,4-glycosidic bond between N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) in the murein sacculus (Figure 1-3). The lysozymes accomplish a similar reaction as endomuramidases by hydrolyzing the ß-1,4-glycosidic bond, but the lytic transglycosylases carry out an intramolecular glycosyl transferase reaction to produce 1,6-anhydromuropeptides. Lytic transglycosylases are implicated in remodeling of the murein during growth and cell division and in the recycling of peptidoglycan material. Three lytic transglycosylases in *E. coli* are encoded by *sltY*, *mltA*, and *mltB* (50). The

structure of the large soluble lytic transglycosylase (Slt70) has been solved, which revealed a catalytic domain with similar features of the lysozymes (58). The smaller lytic transglycosylases encode outer membrane lipoproteins. One of these (MltA) was originally identified as a soluble protein (Slt35), which is now known to be released into the aqueous milieu by proteolysis of the lipoprotein anchor (50).

The proposed mechanism for lysozymes and lytic transglycosylases depends on a catalytic glutamic acid residue that functions as a general acid by donating a proton to the oxygen of the ß-1,4 glycosidic bond leading to the formation of an oxocarbonium ion intermediate (59). In hen egg white lysozyme, oxocarbonium ion formation is stabilized by steric distortion of the MurNAc residue into a sofa conformation and by electrostatic stabilization with a neighboring aspartate residue. The leaving group is then replaced by a water molecule, which is presumably activated by the glutamate for nucleophilic attack on the C1 carbon of the oxocarbonium ion. The reaction proceeds with retention of configuration and results in the production of reducing muropeptides (59).

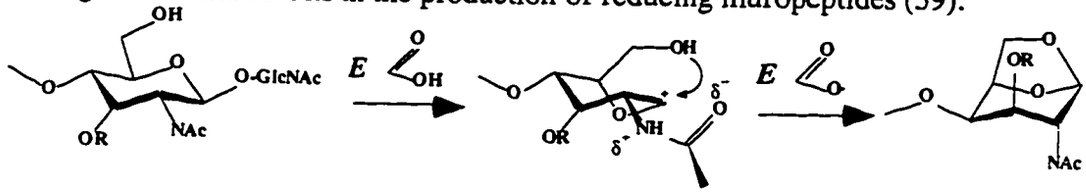


Figure 1-3. Formation of 1,6-anhydromuropeptides by lytic transglycosylases (60).

In Slt70, the C6-hydroxyl group of MurNAc performs an intramolecular nucleophilic attack on C1 of the oxocarbonium ion, which must adopt a boat conformation to orient the C5-hydroxymethyl group into an axial position (60). The glutamate anion presumably activates the C6-hydroxyl group for nucleophilic attack, but the electrostatic aspartate is absent in Slt70. Since the intramolecular reaction does not depend on the diffusion of a water molecule to replace the leaving group, electrostatic stabilization of the oxocarbonium ion intermediate may not be required. However, on the basis of an Slt70 complex with an oxocarbonium ion analog, it was proposed that the 2-acetamido group of MurNAc can participate in stabilization of the boat conformation of

the oxocarbonium ion and possibly even react to form an oxazoline intermediate (60). The resulting non-reducible 1,6-anhydromuropeptides are recycled by the bacterium.

In the process of muropeptide recycling, the amide-linked peptide side chain associated with the D-lactyl ether substituent at O3 of MurNAc can be removed by N-acetylmuramyl-L-alanine amidase. In *E. coli*, a periplasmic amidase encoded by *amiA*, recognizes both reducing and 1,6-anhydromuropeptides to produce a tripeptide, L-alanyl-D-γ-glutamyl-(L)-*meso*-diaminopimelate, which is transported into the cell by the oligopeptide permease and an additional unspecified system (61). Any C-terminal D-ala-D-ala substituents that remain attached to the L-center of *meso*-diaminopimelate of the muropeptide are sequentially removed in the periplasm by the action of DD-carboxypeptidases and an LD-carboxypeptidase (50). Once in the cytoplasm, the tripeptide is coupled to nucleotide activated precursors of peptidoglycan biosynthesis by a UDP-MurNAc::tripeptide ligase encoded by *mpl* (62). However, the primary recycling pathway depends on a 1,6-anhydromuropeptide permease encoded by *ampG*, and a cytoplasmic 1,6-anhydromuropeptide-specific amidase encoded by *ampD* (63). A cytoplasmic ß-N-acetylglucosaminidase can remove the GlcNAc residue to form 1,6-anhydroMurNAc-tripeptide, which is also a substrate for AmpD. Both AmpG and AmpD were originally identified in the regulation of the inducible class C ß-lactamase (AmpC), which led to the discovery that muropeptide recycling can function as a regulatory device that communicates extracellular signals to intracellular transcriptional activators (64). The *ampC* gene can be regulated a the level of transcription by AmpR, which is activated *in vivo* by 1,6-anhydro-MurNAc-tripeptide. This species accumulates in the cytoplasm of *ampD* mutants and is believed to serve as a signal of disturbances in peptidoglycan metabolism brought on by the action of ß-lactam antibiotics in the periplasm.

Lipoproteins of plasma and bacterial cell envelopes

Lipids are transported in the plasma in association with protein-containing particles known as lipoproteins (65). Lipoproteins are classified by their density, which reflects their lipid content. Very low density lipoproteins (VLDL), synthesized by the liver, and chylomicrons, synthesized by the intestine, transport triacylglycerols, phospholipids, cholesterol, and cholesteryl-ester to peripheral tissues. The gradual degradation of VLDL by removal of triacylglycerol leads to the production of low density lipoprotein (LDL), which is removed from circulation by receptor mediated endocytosis. High density lipoproteins (HDL) are produced by the liver without any cholesteryl-esters, but these are acquired by HDL in the circulation.

Plasma lipoproteins share the same basic structure of a hydrophobic lipid core surrounded by a monolayer of phospholipid and cholesterol into which apolipoproteins are embedded. Most apolipoproteins belong to a homologous multigene family composed of repeated amphipathic α-helical segments, but others have distinct structural features (65). Apolipoproteins can have structural, catalytic or regulatory functions in lipid metabolism. For example, lecithin cholesterol acyltransferase is activated by apolipoprotein-AI and transfers fatty acyl groups from phosphatidylcholine (lecithin) to the hydroxyl group of cholesterol in HDL. This reaction was once believed to be assisted by the HDL associated apolipoprotein D, which belongs to a distinct family of lipid-binding proteins known as lipocalins.

The lipocalin family of small extracellular proteins have diverse functions associated with their interactions with hydrophobic ligands (66). Lipocalin functions include lipid transport, coloration, olfaction and pheromone transport, and a catalytic role in prostaglandin biosynthesis. Most lipocalins share conserved sequence motifs, which define the kernel lipocalins, while a smaller group of outlier lipocalins can also be discerned. Despite an intrinsically low level of amino acid sequence identity, the lipocalins share highly conserved three dimensional structures characterized by an 8-

stranded antiparallel ß-barrel followed by a C-terminal α-helix. The ß-strands form two stacked orthogonal sheets which enclose the internal ligand binding site. Several arthropod lipocalins bind biliverdin IXγ, which imparts a blue color in complex with the binding protein (insecticyanin) to provide insect camouflage or photoprotection. Until recently, lipocalins were exclusive to Eukarya, but a bacterial lipocalin (Blc) was recently identified as an outer membrane lipoprotein in *E. coli*. Remarkably, the lipocalin of greatest similarity to Blc was found to be ApoD (36).

Until this discovery, the only relationship between bacterial lipoproteins and plasma lipoproteins was the common name. Bacterial lipoproteins are anchored to the cell envelope by a lipid-modified N-terminal cysteine residue (67). Most bacterial lipoproteins in *E. coli* are targeted to the inner leaflet of the outer membrane by a specific transport mechanism (68). However, some lipoproteins remain associated with the outer leaflet of the cytoplasmic membrane and these usually specify negatively charged residues at the +1 and +2 positions of the mature lipoprotein (69). Lipoproteins in Gram-positive bacteria necessarily reside on the outer surface of the cytoplasmic membrane, but some lipoproteins can be excreted as micellar complexes (70). The discovery of an Archaeal lipoprotein (halocyanin) was surprising because Archaea contain unusual isopranyl lipids, but it indicates that lipoprotein processing machinery must have been present in the last common ancestor (71). However, lipoprotein processing machinery was not transferred to Eukarya. One common feature of plasma apolipoproteins and bacterial lipoproteins is that the inner leaflet of the outer membrane in Gram-negative bacteria and the surface of plasma lipoproteins represent the only two examples in biology of asymmetric bilayer membranes with a single leaflet of phospholipid. The external leaflet of the Gram-negative outer membrane comprises the lipid A component of lipopolysaccharide (72).

Bacterial lipoproteins are specified by an N-terminal type-II signal peptide that can be identified by a conserved motif around the invariant cysteine at the cleavage site.

Like type-I signal peptides, at least one positively charged residue at the N-terminus is followed by a hydrophobic stretch preceding the cleavage site. The type-II signal peptide targets the nascent lipoprotein to the cytoplasmic membrane, where a phosphatidylglycerol::prolipoprotein diacylglyceryl transferase catalyzes the synthesis of a diacylglyceryl prolipoprotein (67). The diglyceride is attached to the cysteine as a thioether, which is unprecedented in biology. The diacylglyceryl prolipoprotein is a substrate for signal peptidase II, which cleaves the signal peptide to expose the α-amino group of the diacylglycerylcysteine residue. Signal peptidase II can be specifically inhibited by an antibiotic known as globomycin. In some bacteria, the thioether linked diglyceride solely anchors the lipoprotein to the cell envelope, but in *E. coli* this is an intermediate known as apolipoprotein (not to be confused with plasma apolipoproteins). A phospholipid::apolipoprotein transacylase can modify the α-amino group with an amide-linked fatty acid to produce a mature lipoprotein (67). In Mycoplasmas, the type-I signal peptidase machinery has been lost and so the cell depends entirely on lipoproteins as a means of protein secretion (73).

Small multidrug resistance proteins and cytochrome oxidases

The smallest known proton-motive efflux pumps belong to the small multidrug resistance (Smr) family (74). These proteins are typically 110 amino acid residues in size with 4 transmembrane spanning segments and have been proposed to act as multimers. Some members are so hydrophobic that they exhibit the properties of proteolipids that can be extracted into 1:1 $CHCl_3$:methanol and reconstituted into liposomes where they remain functional. The best characterized members are plasmid encoded drug resistance proteins, but one chromosomal drug resistance protein has been identified in *E. coli*. Another chromosomal Smr protein encoded by *sugE* has no known substrate, but is conserved in several bacteria including *E. coli*. SugE was identified as a multicopy suppressor of a mutation in *groE*, which encodes a molecular chaperone (75). However,

a sequencing error led to the misidentification of an N-terminal extension known as SugEL and the failure to identify a smaller open reading frame in the *sugEL* extension which was present on the clone used throughout the original investigation. Therefore, the original identification of SugE as a suppressor of GroE may not be valid. However, the close evolutionary relationship of SugE with those Smr members that have been established as proton motive efflux pumps suggests that SugE exerts a similar function with an as yet unspecified substrate.

The proton motive force that energizes drug efflux by the Smr family can be generated under aerobic conditions in part by the action of a cytochrome oxidase. This enzyme contributes to the generation of a proton motive force by coupling reducing equivalents from metabolism to the reduction of molecular oxygen (76). Two distinct strategies have evolved in Bacteria, which can transfer reducing equivalents from the quinol pool to cytochrome C and then to subunit II of cytochrome C oxidases, or to directly transfer reducing equivalents from the quinol pool to the quinol oxidases. In each case, subunit II is anchored to the membrane complex by two N-terminal transmembrane α-helical segments and exposes a large cytosolic domain that exhibits the Greek key-ß-barrel motif of the cupredoxins, which can bind a single copper atom. Like the cupredoxins, subunits II can bind copper, but they have a binuclear copper center in the cytochrome C oxidases while the quinol oxidases have lost copper altogether. Mitochondria were derived from an ancestor of the α-subdivision of proteobacteria, which can possess both cytochrome C and quinol oxidases, but only the cytochrome C oxidase is found in modern mitochondria (77).

Programmed cell death modules

Programmed cell death provides a practical means to regulate the growth and development of multicellular organisms. However, bacteria can also benefit from the programmed death of specific subpopulations under starvation conditions (78). In *E.*

coli, several plasmids are known to be stabilized in populations by a post segregational killing effect. When a plasmid bearing cell loses a plasmid, the cell is killed. The mechanisms have been traced to a group of similarly designed addiction systems. For example, the plasmid emergency maintenance system encoded by *pem* of plasmid R100 is organized as tandem genes that encode an antidote and a toxin. The antidote is intrinsically labile and so the cell is addicted to the continuous production of the antidote. When a cell loses the R100 plasmid, the degradation of the antidote releases the killing action of the toxin (79). This simple killing mechanism ensures that the plasmid is maintained in the population.

A number of plasmid-encoded antidote-toxin gene pairs have been previously ascribed various functions, but are better regarded as addiction modules. For instance, the pairing on plasmids of genes that encode restriction and modification enzymes, which can protect cells from foreign DNA, render the plasmid-bearing cell addicted to the continued production of the modification enzyme. The paradox of rare-cutters, which are unlikely to provide defense against small viral genomes are easily comprehended in terms of addiction (80). Additionally, the colicin and immunity determinants of colicin plasmids were once thought to function primarily as an antibiotic mechanism of cellular defense. However, the narrow spectrum of colicins against cells that are compatible with colicin plasmids can be rationalized in terms of addiction because cells bearing the colicin plasmid are addicted to the continued production of the immunity determinant (81).

Plasmid addiction cannot be regarded as true programmed cell death if one considers that plasmids are parasitic DNA. However, at least one chromosomal addiction module has been identified in association with programmed cell death. The *mazEF* addiction module of *E. coli* is homologous to the *pem* locus, but resides within an operon that includes the upstream *relA* gene (82). The RelA protein is a ribosomal subunit that senses uncharged amino-acyl tRNAs, and contributes to the synthesis of ppGpp as part of

the stringent response. Continuous transcription of the *relA* operon (and *mazEF*) under growth conditions prevents cell death. Under starvation conditions, ppGpp contributes to the repression of *mazEF* transcription, which unleashes the cell killing effect. It has been argued that a moribund subpopulation of starving bacteria can contribute to the survival of the remaining cells in the population if they release their cellular contents by bacteriolysis (82). However, the killing action of *mazEF* is not intrinsically bacteriolytic, but this addiction module provides evidence that true programmed cell death may be prevalent in Bacteria.

Thesis problem

These studies evolved naturally from earlier investigations on the fumarate reductase respiratory complex in the cytoplasmic membrane of *Escherichia coli*. The cloning and sequencing of the *frd* operon in the laboratory of Dr. Stewart Cole in 1982 led to the identification of four subunits in the fumarate reductase complex. However, the AmpC ß-lactamase under investigation in the laboratory of Dr. Staffan Normark led to the discovery that the *ampC* gene was intimately associated with *frd* since the *ampC* promoter resided within the coding sequence of the *frdD* cistron. Therefore, the *ampC* promoter directed transcription through the *frd* Rho-independent terminator, which also functioned as a growth rate-dependent attenuator for *ampC* transcription known as *ampA*. The *ampA* leader sequence contains a ribosome binding site preceding an initiating methionine codon which is followed by a stop codon. The binding of a ribosome to the *ampA* leader is believed to cause anti-termination and transcription of the *ampC* open reading frame. Since ribosome synthesis is coordinated with growth rate by the stringent response, AmpC synthesis is coordinated with growth rate by *ampA*. However, the role of growth rate in the regulation of a ß-lactamase is unclear.

In *Enterobacter cloacae* and *Citrobacter freundii*, which are closely related to *E. coli* as members of the *Enterobacteriaceae*, the *frd* and *ampC* operons were found to be

separated by a distinct regulatory gene known as *ampR*. The presence of *ampR* rendered the divergently oriented *ampC* inducible by ß-lactam antibiotics, but the *E. coli ampC* operon remained non-inducible even in the presence of *ampR*. However, the cloned *ampRC* locus was inducible when carried in *E. coli* and unlinked mutations that rendered the inducible *ampRC* locus constitutively induced in *E. coli* were identified as *ampD*. Drs. Cole and Normark independently and simultaneously published the sequence of the cloned *ampD* gene from *E. coli* and discovered that it was part of an operon that included *ampE* (83,84). Since AmpD was a cytoplasmic protein and AmpE was located in the cytoplasmic membrane, it was proposed that the *ampDE* operon specified a signal transduction pathway involved in ß-lactamase induction. Dr. Weiner renewed his collaboration with Dr. Cole by investigating the biochemical properties of the membrane component AmpE.

I embarked on my graduate studies with the objective of elucidating the signal transduction pathway in ß-lactamase induction. At that time, a series of investigations on signal transduction in bacteria had identified a family of so-called two-component regulatory systems. Although *ampDE* was evolutionarily distinct from any known regulatory system, we suspected that it may present a variation on the theme of two-component regulation. However, I was perplexed by a number of relationships between the inducible and growth-rate dependent forms of *ampC* regulation. For instance, why did *E. coli* carry *ampDE* when it had lost *ampR*, which was essential for induction? Perhaps *ampDE* had an additional function independent of ß-lactamase regulation. Why did *E. coli* utilize the *ampA* attenuator, which produced insufficient ß-lactamase to render cells clinically resistant to antibiotics in the absence of mutational up-regulation? Perhaps *ampC* also had an additional physiological role. I believed that the elucidation of the *ampC* induction mechanism hinged on the complete understanding of AmpC function.

In Chapter 2, I present a hypothesis that I published after carefully considering the literature at that time. I was influenced mainly by recent developments in the analysis of the murein sacculus from the laboratory of Dr. Joachim Volker-Höltje at the Max Planck Institute for Developmental Biology in Tübingen Germany. Prior to the application of high performance liquid chromatography to the analysis of muramidase digests of murein, the murein sacculus was believed to be composed of a few major components connected by peptide bridges between D-alanine and the D-center of *meso*-diaminopimelate. However, Dr. Höltje's lab discovered a complexity that included a minor class of peptide bridges connected directly between the L- and D-centers of neighboring diaminopimelate residues. Remarkably, the levels of the minor LD-peptide bridges were found to be dependent on growth rate in a manner inversely proportional to the expression of the *E. coli ampC* gene.

Since a structural similarity between penicillin and acyl-D-alanyl-D-alanine had successfully predicted the mechanism of penicillin action in 1965, I suspected that the LD-peptide bridges should also share a structural similarity with penicillin, which turned out to be just as extensive as that seen between penicillin and acyl-D-alanyl-D-alanine. Therefore, I proposed that AmpC may possess an additional biochemical activity as an LD-endopeptidase in addition to its established function as a ß-lactamase. Since many inducible enzyme systems depend on regulatory metabolites derived from the metabolic pathway in which the enzyme participates, I suspected that AmpC induction may depend on a murein metabolite. The rudiments of a peptidoglycan recycling pathway had been uncovered at that time and so I proposed a corollary that *ampC* induction may depend on a recyclable cell-wall peptide.

This hypothesis laid the foundation for several years of experimentation. I initiated my work by focusing on AmpR, which belonged to the LysR family of transcriptional regulators. Since AmpR binds to its own promoter as a repressor, I had to devise a method to clone the gene for overexpression. The polymerase chain reaction

(PCR) had been introduced into the biochemistry department at that time, and so I was the first to apply PCR in the Weiner lab for the purpose of overexpressing AmpR. However, I soon learned that AmpR production was limited by a poor ribosome-binding site, and so I adapted PCR mutagenesis in order to introduce a good ribosome binding site into the *ampR* gene. Massive AmpR production was achieved, and this PCR mutagenesis method has been utilized successfully in the analysis of many different proteins in the Weiner lab. However, the AmpR protein was overexpressed as an insoluble aggregate.

I undertook a biochemical investigation of the solubility properties of AmpR, which culminated in a purification procedure described in Chapter 3 that involved differential precipitation and gel-filtration chromatography. The purified AmpR was active in site-specific DNA binding and existed in solution as a homodimer, which was expected of a helix-turn-helix DNA-binding protein. However, the *ampR* operator determined earlier by Nadine Honoré and Stewart Cole did not exhibit the characteristic inverted repeat signature of other helix-turn-helix DNA-binding proteins. I utilized purified AmpR to demonstrate that AmpR interacts with its operator as a pair of dimers in close contact and showed that an inverted repeat was discernible in one half of the operator. Presumably, a recyclable cell wall peptide acted as a second messenger of ß-lactam induction by serving as an allosteric activator for AmpR. In order to test this idea, I would need to learn the technology of murein analysis.

I established a collaboration with Dr. Höltje and applied for an Annual Grant from the German Academic Exchange Service so that I could spend a year in Dr. Höltje's laboratory as an exchange student. This grant was eventually awarded, but in the mean time I investigated the possibility that AmpC may have an additional physiological function. A laboratory deletion mutant (*E. coli* MI1443) in the *frd* operon was also deleted in *ampC*. Therefore, I investigated whether *ampC* could affect the phenotypic properties of this mutant. I utilized the gene encoding the class A RTEM ß-lactamase as

a control since RTEM is an AmpC homolog, but it was not expected to have any function beyond that of a ß-lactamase. Additionally, the laboratory of Dr. Michael James was actively pursuing crystallographic studies of purified RTEM and AmpC, and kindly provided samples of their purified enzymes that were useful for my analyses. In Chapter 4, I present the discovery of a growth defect in *E. coli* MI1443 that was specifically complemented by the expression of AmpC. While these findings were attributed to either an effect of AmpC expression or to the ß-lactamase activity of AmpC, I had little doubt in my mind that ß-lactamase activity would prove to be associated with LD-endopeptidase activity, but proof of this would also have to await my arrival in Germany.

I continued my investigation by creating an active site mutant of AmpC in which the catalytic serine residue was converted to an alanine. When I characterized the effect of this AmpC mutant in *E. coli* MI1443, I was surprised to find that the mutant was complemented to the same extent as the wild type AmpC. I began to have my first doubts that AmpC would prove to be an LD-endopeptidase. At that time, Dr. Gary Cecchini's lab described a different *frd* deletion, which was also defective in *ampC* as well as the upstream *genX*, but was less extensive than the deletion in *E. coli* MI1443. I obtained the mutant from Dr. Cecchini and was encouraged to find a growth defect under aerobic conditions. However, this defect could not be complemented by plasmids that carried either the *C. freundii ampRC* or the *E. coli ampAC* loci, but was complemented by *genX*, which was transcribed divergently from *frd*. I now began to have serious doubts that AmpC would prove to be an LD-endopeptidase. However, I noticed that Dr. Cecchini's strains, which were derived from *E. coli* MC4100, underwent a stationary phase bacteriolysis that was traced to an uncharacterized region downstream of *ampC*.

I was fascinated by this finding and discussed it with Dr. Laura Frost, who indicated that MC4100 had special stationary phase properties because it was wild type in a gene called *katF*. This gene was cloned by Dr. Michael Mulvey as a graduate student who happened to be a postdoctoral fellow in the biochemistry department at that time. I

soon became aware that *katF* encoded a stationary phase-specific sigma factor that was re-named RpoS by Dr. Regine Hengge-Aronis, who provided me with an *rpoS* mutant in an MC4100 background. I soon discovered that stationary phase bacteriolysis was RpoS dependent and that a gene downstream of *ampC* described previously as *sugE* was probably responsible since SugE expression was reported to be induced in stationary phase. I also sequenced the gap between *sugE* and *ampC* to uncover an additional gene that shared features with the lipocalin protein family and which carried a type-II signal peptide. Shortly before I embarked for Germany, I established a collaboration with Dr. Frost and her graduate student Sonya Penfold, who was exploring RNA transcription in *E. coli*. The goal was to demonstrate that *sugE* transcription was *rpoS*-dependent by utilizing primer extension with RNA from *E. coli* MC4100 and the *rpoS*-derivative. For good measure, I suggested that the lipocalin gene I had uncovered could be investigated as a negative control.

After I arrived in Dr. Höltje's lab in 1994, he presented evidence obtained in collaboration with Dr. Bernd Wiedemann, which demonstrated that purified AmpD functioned as a 1,6-anhydro-N-acetylmuramyl-L-alanine amidase. Dr. Höltje discovered the lytic transglycosylase when he was a graduate student and was keenly aware of the 1,6-anhydro bond in *E. coli* muropeptides. His findings strongly buoyed my enthusiasm that AmpC may still prove to be an LD-endopeptidase after all, and so my investigations began. However, all evidence indicated that AmpC was not an LD-endopeptidase, nor was it likely involved in the metabolism of 1,6-anhydro muropeptides. Dr. Höltje then returned from a Gordon Conference where he had attended a seminar by Dr. Normark who presented evidence that 1,6-anhydromuropeptides were recycled by *E. coli* and were involved in the regulation of *ampC* induction by a pathway very similar to the one I had proposed earlier. I was pleased to learn that the two aspects of my hypothesis had been investigated and that it had been useful. However, I also realized that I would be wasting my time to continue the investigation of *ampC* regulation while I was in Germany. The

AmpR purification procedure I had developed was eventually instrumental in the proof by Dr. Normark's lab that AmpR is indeed regulated by a recyclable cell wall peptide. Moreover, a D-stereospecific endopeptidase with ß-lactamase activity was isolated from *Bacillus cereus* and shown to be homologous to class C ß-lactamases. My results demonstrating that AmpC is not an LD-endopeptidase are presented in Chapter 5.

Investigations in Dr. Höltje's laboratory before my arrival had uncovered two genes for additional lytic transglycosylases in *E. coli*. One of these was found to possess a type-II signal peptide and a graduate student named Kerstin Ehlert was preparing a lipoprotein characterization. I decided I could spend the remainder of my time in Germany by performing a lipoprotein characterization on the lipocalin gene, which I had named *blc*. Kerstin showed me techniques that Dr. Höltje picked up when he was a post doctoral fellow with Dr. Volkmar Braun, who had discovered the first bacterial lipoprotein. I also began to receive reports from Dr. Frost that *sugE* was not *rpoS*-dependent but that *blc* actually was. Since Blc expression did not result in stationary phase bacteriolysis, I suspected that the samples might have been mixed up. The experiment was repeated but the same result was reported. Dr. Raymond Turner in the Weiner lab had taken up studies of *sugE* and had not reported any observations of stationary phase bacteriolysis. Therefore, if *blc* and *sugE* were not responsible for stationary phase bacteriolysis, what was? The answer would have to wait until I returned to Canada.

On my return, I confirmed that *blc* transcription was *rpoS*-dependent by creating a translational gene-fusion to *lacZ* and monitoring ß-galactosidase activity. The fact that *blc* was *rpoS*-dependent and also a globomycin sensitive outer membrane lipoprotein raised the possibility that Blc might contribute to the general state of resistance of stationary phase cells to environmental stress. The characterization of Blc is presented in Chapter 6. Perhaps the most surprising feature of Blc is its evolutionary relationship with the lipocalins. In Chapter 7, I propose that Blc belongs to an ancient group of lipocalins

that are membrane localized, and that modern soluble lipocalins must have resulted from a loss of the ancestral membrane anchor. I also present findings that Blc production is associated with a blue chromophore that is chloroform extractable, and that the *blc* promoter belongs to a subclass of *rpoS*-dependent promoters that are expressed optimally under high osmolarity conditions. Additionally, I demonstrate that subunits II of the cytochrome oxidases have an ancient ancestry as soluble proteins that eventually became membrane localized, probably as lipoproteins. Studies of *E. coli cyoA* were performed with a Cys25Ala mutant provided by Dr. Robert Gennis.

Returning to my interest in stationary phase bacteriolysis, I extended my sequence analysis beyond *sugE* in *C. freundii*. The homologous region was now available for *E. coli* from the MG1655 genome project, which had suggested that a sequencing error had been made in the original analysis of *sugE*. This had been characterized from plasmid pLC16-43 of the Clarke and Carbon gene bank of *E. coli* CS520. The Weiner lab had utilized the same clone to characterize *frd*, and so I was able to confirm that a sequencing error had indeed been made in *sugE* and showed that the *C. freundii sugE* was homologous. However, an additional regulatory gene, which belonged to the OmpR family of transcriptional regulators, was transcribed divergently from *sugE* in *C. freundii*. Beyond the response regulator was a gap of 450 bp before the Rho-independent terminator of *efp*, which was the final truncated open reading frame on the clone. A similar gap was present between *efp* and *sugE* in *E. coli*, but this is where the sequencing error had been made. Since a promoter had been reported upstream of *sugE* in *E. coli*, I made a close inspection for small open reading frames, since none had been reported by the genome project.

To my surprise, I saw type-II signal peptides preceding each of a pair of open reading frames, which were both less than 50 codons in size. Remarkably, they were conserved in both *E. coli* and *C. freundii*, and they were paralogous within each organism. Furthermore, the upstream promoter from the *sugE* analysis preceded the

second lipoprotein gene, which was followed by a Rho-independent terminator. The first lipoprotein gene was located immediately downstream of the *efp* Rho-independent terminator and the cloned genes were responsible for stationary phase bacteriolysis. In order to confirm the bacteriolytic effect, I expressed the tandem genes from an inducible promoter and was surprised to find that the lytic effect was actually reduced. I then split the genes apart and found that the second gene was responsible for bacteriolysis and the first gene served as an antidote. The lipoprotein nature of the two gene products was confirmed and I named them entericidins A and B, because they seemed to be involved in enterobacterial suicide.

In order to confirm that entericidin B was governed by an RpoS-dependent promoter, I returned to Dr. Frost who suggested I work in the neighboring lab with Dr. Brenda Leskiw who was already in the process of studying an RNA molecule. On a hunch, I suspected that the response regulator in *C. freundii* might be replaced by the EnvZ/OmpR two-component regulatory system in *E. coli* and I obtained *envZ* and *ompR* mutants from Dr. Tom Silhavy. In Dr. Leskiw's lab, I demonstrated that the entericidin B promoter was indeed a member of the subclass of *rpoS*-dependent promoters that operate optimally under high osmolarity conditions and that EnvZ/OmpR repressed the entericidin B promoter in response to high osmolarity. Therefore, high osmolarity had both positive and negative effects on the same promoter. The tight regulation of the entericidin locus may reflect a role in programmed cell death.

Since programmed cell death in bacteria has been proposed to share features with antibiotic-induced death, we were encouraged to find that the entericidins exhibited amphipathic α-helical properties characteristic of antimicrobial peptides. In collaboration with Dr. Robert Hodges and Dr. Cyril Kay, synthetic peptides in which the N-terminal tripalmitoylcysteine residues of the entericidins were replaced with N-acetylserine were characterized by circular dichroism. In the presence of a membrane mimetic solvent, α-helical conformation was confirmed, suggesting that the membrane bound entericidins

may indeed regulate membrane stability as amphipathic α-helices. The structural basis for the differential effects of the paralogous entericidins remains to be determined. However, the entericidin locus provides a weak link in the armor of stationary phase bacteria, which are intrinsically resistant to ß-lactam antibiotics. The entericidin discovery resulted unexpectedly from an investigation of the growth rate dependent regulation of the AmpC ß-lactamase and may be useful in the design of stationary-phase specific antimicrobial agents.

References:

(1) Davies, J. (1994) *Science*, **264:** 375

(2) Woese C.J. (1994) *Microbiol Rev.*, **58:** 1

(3) Hartl, D.L., Medhora, M., Green, L., and Dykhuizen, D.E. (1986) *Phil. Trans. R. Soc. Lond.* B **312:** 191

(4) F.C. Neidhardt, ed. (1987) *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, ASM Press, Washington, D.C.

(5) F.C. Neidhardt, ed. (1996) *Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition,* ASM Press, Washington, D.C.

(6) Fristrom, J.W., and Spieth P.T. (1980) *Principles of Genetics*, Chiron Press inc., New York.

(7) Dawkins, R. (1982) *The Extended Phenotype*, Oxford University Press, Oxford, U.K.

(8) Haeckel, E . (1866) *Generelle Morphologie der Organismen*, Reimer, Berlin.

(9) Whittaker, R.H. (1959) *Q. Rev. Biol.* **34:** 210

(10) Chatton, E. (1938) *Titres et Travoux Scientifiques (1906-1937) de Edouard Chatton*, E. Sottano, Séte, France.

(11) Woese, C.R. (1987) *Microbiol. Rev.* **51:** 221

(12) Bernal, J.D. (1962) *in Horizons in Biochemistry*, Academic Press, New York.

(13) Zuckerkandl, E., and Pauling, L. (1965) *J. Theoret. Biol.* **8:** 357

(14) Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. USA* **87:** 4576

(15) Morell, V. (1996) *Science* **273:** 1043

(16) Morell, V. (1997) *Science* **276:** 699

(17) Pace, N.R. (1997) *Science* **276:** 734

(18) Reeck, G.R., et al. (1987) *Cell* **50:** 667

(19) Fitch W.M. (1970) *Syst. Zool.* **19:** 99

(20) Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86:** 9355

(21) Doolittle, W.F., and Brown, J.R. (1994) *Proc. Natl. Acad. Sci. USA* **91:** 6721

(22) Baldauf, S.L., Palmer, J.D., and Doolittle, W.F. (1996) *Proc. Natl. Acad. Sci. USA* **93:** 7749

(23) Margulis, L. (1993) *Symbiosis in Cell Evolution*, Freeman, New York, 2nd Ed.

(24) Germot, A., Philippe, H., and Le Guyader, H. (1996) *Proc. Natl. Acad. Sci. USA* **93:** 14614

(25) Roger, A.J., Clark, C.G., and Doolittle, W.F. (1996) *Proc. Natl. Acad. Sci. USA* **93:** 14618

(26) Pace, N.R. (1991) *Cell* **65:** 531

(27) Selander, R.K., Caugant, D.A., and Whittam, T.S. (1987) Genetic Structure and Variation in Natural Populations of *Escherichia coli*, *in Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(28) Ochman, H., and Wilson, A.C. (1987) Evolutionary History of Enteric Bacteria, *in Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(29) Brenner, D.J. (1984) Family 1 *Enterobacteriaceae* Rahn 1937, *in Bergey's Manual of Systematic Bacteriology, Vol. 1*, N.R. Krieg and J.G. Holt eds., The William and Wilkins Co., Baltimore.

(30) Grundström, T., and Jaurin, B. (1982) *Proc. Natl. Acad. Sci. USA* **79:** 1111

(31) Jaurin, B., Grundström, T., Edlund, T., and Normark, S. (1981) *Nature* **290:** 221

(32) Lindberg, F., Westman, L., and Normark, S. (1985) *Proc. Natl. Acad. Sci. USA* **82:** 4620

(33) Lindquist, S., Lindberg, F., and Normark, S. (1989) *J. Bacteriol.* **171:** 3746

(34) Lindberg, F., Lindquist, S., and Normark, S. (1987) *J. Bacteriol.* **169:** 1923

(35) Lindquist, S., Weston-Hafer, K., Schmidt, H., Pul, C., Korfmann, G., Erickson, J., Sanders, C., Martin, H., and Normark, S. (1993) *Molec. Microbiol.* **9:** 703

(36) Bishop, R.E., Penfold, S.S., Frost, L.S., Höltje, J.-V., and Weiner, J.H. (1995) *J. Biol. Chem.* **270:** 23097

(37) Paulsen, I.T., Skurray, R.A., Tam, R., Saier Jr., M.H., Turner, R.J., Weiner, J.H., Goldberg, E.B., and Grinius, L.L. (1996) *Molec. Microbiol.* **19:** 1167

(38) Pratt, L.A., and Silhavy, T.J. (1995) Porin Regulon of *Escherichia coli*, *in Two-Component Signal Transduction*, J.A. Hoch and T.J. Silhavy eds., ASM Press, Washington, D.C.

(39) Aoki, H., Adams, S.L., Chung, D.G., Yaguchi, M., Chuang, S.E., and Ganoza, M.C. (1991) *Nucleic Acids Res.* **19:** 6215

(40) Bergström, S., Lindberg, F.P., Olsson, O., and Normark, S. (1983) *J. Bacteriol.* **155:** 1297

(41) Malhotra, A., Severinova, E., and Darst, S.A. (1996) *Cell* **87:** 127

(42) Record Jr., M.T., Reznikoff, W.S., Craig, M.L., McQuade, K.L., and Schlax, P.J. (1996) *Escherichia coli* RNA Polymerase ($E\sigma^{70}$): Promoters and the Kinetics of the Steps of Transcription Initiation, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(43) Hengge-Aronis, R. (1996) *Molec. Microbiol.* **21:** 887

(44) Espinosa-Urgel, M., Chamizo, C., and Tormo, A. (1996) *Molec. Microbiol.* **21:** 657

(45) Hengge-Aronis, R. (1996) Regulation of Gene Expression During Entry into Stationary Phase, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition,* ASM Press, Washington, D.C.

(46) Schell, M.A. (1993) *Ann. Rev. Microbiol.* **47:** 597

(47) Steitz, T.A. (1990) *Quart. Rev. Biophys.* **23:** 205

(48) Rampersaud, A., Harlocker, S.L., and Inouye, M. (1994) *J. Biol. Chem.* **269:** 12559

(49) Ghuysen, J.-M. (1991) *Ann. Rev. Microbiol.* **45:** 37

(50) Höltje, J.-V. (1995) *Arch. Microbiol.* **164:** 243

(51) Asano, Y., Kato, Y., Yamada, A., and Kondo, K. (1992) *Biochemistry* **31:** 2316

(52) Oefner, C., D'Arcy, A., Daly, J.J., Gubernator, K., Charnas, R.L., Heinze, I., Hubschwerlen, C., and Winkler, F.K. (1990) *Nature* **343:** 284

(53) Strynadka, N.C.J., Adachi, H., Jensen, S.E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., and James, M.N.G. (1992) *Nature* **359:** 700

(54) Dubus, A., Normark, S., Kania, M., and Page, M.G.P. (1994) *Biochemistry* **33:** 8577

(55) Labkovsky, E., Moews, P.C., Liu, H., Zhao, H., Frère, J.-M., Knox, J.R. (1993) *Proc. Natl. Acad. Sci. USA* **90:** 11257

(56) Damblon, C., Raquet, X., Lian, L.-Y., Lamotte-Brasseur, J., Fonze, E., Charlier, P., Roberts, G.C.K., and Frère, J.-M. (1996) *Proc. Natl. Acad. Sci. USA* **93:** 1747

(57) Paetzel, M., and Dalby, R.E. (1997) *Trends Biochem. Sci.* **22:** 28

(58) Thunnissen, A.M.W.H., Dijkstra, A., Rozeboom, H.J., Kalk, K., Engel, H., Keck, W., and Dijkstra, B.W. (1994) *Nature* **367:** 750

(59) Strynadka, N.C.J., and James, M.N.G. (1991) *J. Mol. Biol.* **220:** 440

(60) Thunnissen, A.M.W.H., Rozeboom, H.J., Kalk, K.-H., and Dijkstra, B.W. (1995) *Biochemistry* **34:** 12729

(61) Park, J.T. (1996) The Murein Sacculus, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(62) Mengin-Lecreulx, D., van Heijenoort, J., and Park, J.T. (1996) *J. Bacteriol.* **178:** 5347

(63) Jacobs, C., Huang, L.J., Bartowsky, E., Normark, S., and Park, J.T. (1994) *EMBO J.* **13:** 4684

(64) Jacobs, C., Frère, J.-M., and Normark, S. (1997) *Cell* **88:** 823

(65) Li, W.-H., Tanimura, C., Luo, S., Datta, S., and Chan, L. (1988) *J. Lipid Res.* **29:** 245

(66) Flower, D.R. (1996) *Biochem. J.* **318:** 1

(67) Sankaran, K., Gupta, S.D., and Wu, H.C. (1995) *Meth. Enzymol.* **250:** 683

(68) Gennity, J.M., and Inouye, M. (1991) *J. Biol. Chem.* **266:** 16458

(69) Matsuyama, S., Tajima, T., and Tokuda, H. (1995) *EMBO J.* **14:** 3365

(70) Braun, V., and Wu, H.C. (1993) *in Bacterial Cell Wall*, J.-M. Ghuysen and R. Hakenbeck, eds., New Comprehensive Biochemistry, Vol. 27, p. 319, Elsevier, Amsterdam

(71) Mattar, S., Scharf, B., Kent, S.B.H., Rodewald, K., Oesterhelt, D., and Engelhard, M. (1994) *J. Biol. Chem.* **269:** 14939

(72) Raetz, C.R.H (1996) Bacterial Lipopolysaccharides: a remarkable family of bioactive macroamphiphiles, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition,* ASM Press, Washington, D.C.

(73) Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B., and Hermann, R. (1997) *Nucleic Acids Res.* **25:** 701

(74) Paulsen, I.T., Skurray, R.A., Tam, R., Saier Jr., M.H., Turner, R.J., Weiner, J.H., Goldberg, E.B., and Grinius, L.L. (1996) *Molec. Microbiol.* **19:** 1167

(75) Greener, T., Govezensky, D., and Zamir, A. (1993) *EMBO J.* **12:** 889

(76) Castresana, J., Lübben, M., Saraste, M., and Higgins, D.G. (1994) *EMBO J.* **13:** 2516

(77) Castresana, J., and Saraste, M. (1995) *Trends Biochem. Sci.* **20:** 443

(78) Yarmolinsky, M.B. (1995) *Science* **267:** 836

(79) Jensen, R.B., and Gerdes, K. (1995) *Molec. Microbiol.* **17:** 205

(80) Naito, T., Kusano, K., and Kobayashi, I. (1995) *Science* **267:** 897

(81) Helinski, D.R., Toukdarian, A.E., and Novick, R.P. (1996) Replication Control and Other Stable Maintenance Mechanisms of Plasmids, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(82) Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996) *Proc. Natl. Acad. Sci. USA* **93:** 6059

(83) Lindquist, S., Galleni, M., and Normark, S. (1989) *Molec. Microbiol.* **3:** 1091

(84) Honoré, N., Nicolas, M.-H., and Cole, S.T. (1989) *Molec. Microbiol.* **3:** 1121

CHAPTER 2

Coordinate Regulation of Murein Peptidase Activity and AmpC ß-Lactamase Synthesis in *Escherichia coli*[1]

Contrary to what I once thought, scientific progress did not consist simply in observing, in accumulating experimental facts and drawing up a theory from them. It began with the invention of a possible world, or a fragment thereof, which was then compared by experimentation with the real world.

François Jacob

[1]A version of this chapter has been published. Bishop, R.E., and Weiner, J.H. (1992) *FEBS Lett.* **304:** 103

1. INTRODUCTION

The cell-wall of *Escherichia coli* is composed of an outer membrane and a rigid exoskeleton called the murein sacculus which determines cell shape and protects the bacterium from osmotic lysis. The murein layer lies within the periplasmic space which separates the outer membrane from the inner cytoplasmic membrane. Since the murein is continuously remodeled as the bacterium grows and divides, its synthesis is tightly coordinated with the cell division cycle. Additionally, since the murein is both essential and unique to bacteria, the enzymes of murein metabolism are important targets of antibiotic action. Penicillins and related ß-lactam antibiotics are selective inhibitors of the penicillin-binding proteins (PBPs) which are enzymes of murein biosynthesis (reviewed in [1]).

The PBPs construct murein from peptidoglycan precursors which are synthesized in the cytoplasm. The structure of the peptidoglycan that becomes incorporated into the murein layer of *Escherichia coli* is shown in figure 2-1. Glycan chains of alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) are linked by ß-1,4-glycosidic bonds. The glycan chains are interconnected by short peptides which are attached by amide bonds to the carboxyl group of each MurNAc residue. In enterobacteria, the cell-wall peptides are derived from the pentapeptide L-alanyl-D-γ-glutamyl-(L)-*meso*-diaminopimelyl-D-alanyl-D-alanine [L-ala-D-γ-glu-(L)-*m*-A_2pm-D-ala-D-ala] [1].

Soon after peptidoglycan is incorporated into the glycan network of the murein, the terminal D-ala-D-ala peptide bond is cleaved by a DD-peptidase. The enzyme catalyzes nucleophilic displacement between an active site serine hydroxyl group and the carbonyl carbon of the peptide to form an intermediate acyl-enzyme. The carbonyl carbon of the acyl-enzyme ester bond then reacts with an acceptor nucleophile which regenerates the active serine enzyme and releases the modified peptide. The DD-

transpeptidases employ the ε-amino group of a neighboring *m*-A_2pm residue as acceptor to form a D-ala-(D)-*m*-A_2pm cross-link which interconnects the glycan chains of the murein sacculus (figure 2-1). Approximately one quarter of the cell-wall peptides in *Escherichia coli* are cross-linked by transpeptidation [1]. The DD-carboxypeptidases employ water as acceptor and thereby convert the remaining pentapeptides to tetrapeptides; the D-ala-(D)-*m*-A_2pm cross-link is structurally similar to D-ala-D-ala and subject to hydrolysis by a DD-endopeptidase. These murein hydrolases are necessary for cell growth and division since they help maintain the dynamic structure of the murein sacculus (reviewed in [2]). The family of DD-peptidases are inactivated on binding penicillin.

The molecular basis of penicillin action was predicted by Tipper and Strominger in 1965; they observed that penicillin is a structural analog of the D-ala-D-ala terminus of cell-wall peptides [3]. Indeed, if the reactive amide bond in the ß-lactam ring of penicillin is superposed on the scissile peptide bond of acyl-D-ala-D-ala, the α-carbon backbone of the peptide can adopt a conformation that is nearly isosteric with penicillin. Tipper and Strominger suggested that transpeptidation might proceed by an acyl-enzyme mechanism which conserves the D-ala-D-ala bond energy needed to form a peptide cross-link in the extracytoplasmic milieu [3]. They also proposed that penicillin may exert its antibiotic properties by acylating the active site of a DD-peptidase to form an inactive penicilloyl-enzyme [3]. An important corollary was that ß-lactamases, which hydrolyze the ß-lactam amide bond, may be derived from the DD-peptidases [3].

It has since been demonstrated that DD-peptidases and serine ß-lactamases catalyze amide bond cleavage via an acyl-enzyme mechanism which involves an active site serine residue located at the junction of two domains (reviewed in [4]). In the presence of penicillin, the DD-peptidases form a stable penicilloyl-enzyme whereas the serine ß-lactamases form a penicilloyl-enzyme which is hydrolytically labile. Since DD-peptidases and serine ß-lactamases of class A and C exhibit similar tertiary structures, it

is now believed that they evolved divergently from an ancestral enzyme [5]. However, the serine ß-lactamases of class A are distinct from the class C enzymes (AmpC) [5]. Some relationships between peptidoglycan structure and the regulation of enterobacterial AmpC synthesis are difficult to interpret in terms of the DD-peptidase analogy. This article will describe these relationships and interpret them by analogy with LD-peptidases.

2. METABOLISM AND PHYSIOLOGY OF LD-PEPTIDES

As the murein matures, the (L)-*m*-A_2pm-D-ala peptide bond in cell-wall tetrapeptides can be ruptured by several LD-peptidases. The LD-carboxypeptidase of *E. coli* is insensitive to penicillin though it is inhibited by nocardicin A, a monocyclic ß-lactam which interacts with the enzyme non-covalently [6]. It has been proposed that D-amino acids and a related substituent in nocardicin A are important for LD-carboxypeptidase inhibition [7]. No mutants in the LD-carboxypeptidase gene are available nor has the enzyme been characterized structurally. However, it is probable that the LD-carboxypeptidase operates by an acyl-enzyme mechanism analogous to that of the DD-carboxypeptidases. The LD-carboxypeptidase produces murein tripeptides which are preferentially used as acceptors by PBP3, a DD-transpeptidase required for septation [8]. It is possible that the LD-carboxypeptidase is coordinated with septation since its expression is elevated at the time of cell division [9].

The (L)-*m*-A_2pm-D-ala peptide bond is also important for attachment of Braun's lipoprotein to the murein sacculus. Approximately one third of lipoprotein molecules in *E. coli* are covalently attached through the ε-amino group of their C-terminal lysine to the (L)-*m*-A_2pm residue of a murein tripeptide [10]. By analogy with DD-transpeptidases, the putative (L)-*m*-A_2pm:lipoprotein transpeptidase may catalyze nucleophilic displacement of the (L)-*m*-A_2pm-D-ala peptide bond to form an acyl-enzyme which

employs the lipoprotein as an acceptor. In mutants of *E. coli* and *Salmonella typhimurium* with decreased levels of murein-bound lipoprotein, the outer membrane fails to invaginate with the murein layer during septation [11,12]. However, a mutant of *E. coli* that lacks the lipoprotein grows and divides normally; the abnormally leaky outer-membrane of this mutant suggests the lipoprotein helps maintain the integrity of the cell-wall [13].

Recently, a novel (L)-*m*-A_2pm-(D)-*m*-A_2pm peptide that comprises a small fraction of murein cross-links was discovered in *E. coli*. It has been suggested that an LD-transpeptidase may catalyze nucleophilic displacement of the (L)-*m*-A_2pm-D-ala peptide to form an acyl-enzyme which employs the ε-amino group of a neighboring *m*-A_2pm residue as an acceptor [14]. It has also been suggested that the LD-transpeptidase and the (L)-*m*-A_2pm:lipoprotein transpeptidase may be identical since the lipoprotein is initially attached preferentially to cell-wall peptides in which the *m*-A_2pm residue participates in a cross-link [15]; the (L)-*m*-A_2pm-(D)-*m*-A_2pm cross-links that carry the lipoprotein are twice as abundant as the lipoprotein-free species [16]. Additionally, the (L)-*m*-A_2pm-(D)-*m*-A_2pm peptide bond, and the murein lipoprotein, are the only murein structures regulated by bacterial growth rate; their abundance decreases coordinately with increasing growth rate [17]. This may reflect the activity of an LD-endopeptidase which could modulate the availability of (L)-*m*-A_2pm-(D)-*m*-A_2pm peptides for lipoprotein attachment and liberate lipoprotein-tripeptides from the cross-linked species.

3. GROWTH RATE CONTROL OF AmpC SYNTHESIS AND MUREIN STRUCTURE

The AmpC ß-lactamase of *E. coli* is specifically expressed in direct proportion with growth rate [18]. The chromosomal *ampC* gene is controlled by a promoter which lies within the coding sequence of the neighboring and overlapping fumarate reductase

operon (*frd*) [19]. Transcription of the *ampC* gene is initiated through the *frd* terminator, also known as *ampA*, which functions as an *ampC* attenuator. Antitermination is occasionally achieved when the translational apparatus couples with a ribosome binding sequence in nascent *ampC* transcripts [20]. Since ribosome biogenesis is independently coordinated with growth rate by the Stringent Response, AmpC expression is coordinated with growth rate by the *ampA* attenuator [20]. The *ampC* gene is not induced by ß-lactam compounds and, in the absence of rare mutations that lead to AmpC overproduction, is not normally an agent of ß-lactam resistance [21,22].

The role of growth rate in control of *ampC* gene expression remains an enigma. However, the phenomenon of phenotypic tolerance, whereby slowly growing bacteria are more resistant to the killing action of penicillin, may partly result from structural changes in the cell-wall [23]. The *ampA* attenuator provides levels of AmpC protein that are exactly inversely proportional, over similar growth rates, to levels of both the (L)-*m*-A_2pm-(D)-*m*-A_2pm peptide and the lipoprotein (compare [17] and [18]). To the same degree, cell diameter in *E. coli* is also inversely proportional to growth rate [24]. In *Salmonella typhimurium*, which lacks an *ampC* gene, no correlation between growth rate and cell diameter is evident [25]. The coordinate control of *ampC* gene expression and murein composition in *E. coli* could be explained if the AmpC protein functions as the putative LD-endopeptidase.

4. STRUCTURAL SIMILARITIES BETWEEN CELL-WALL PEPTIDES

Is it possible that the AmpC protein functions both as a ß-lactamase and an LD-peptidase? As demonstrated in figure 2-2, the (L)-*m*-A_2pm-(D)-ala peptide is a structural analog of the D-ala-D-ala peptide and of penicillin. Due to the D-configuration of the ε-carbon in *m*-A_2pm, the two endopeptide cross-links are analogous to their corresponding exopeptides. Therefore, the DD-peptides and the LD-peptides can be largely superposed

on each other and on penicillin. This relationship suggests that DD-peptidases, LD-peptidases and serine ß-lactamases catalyze the cleavage of similar amide substrates and may thereby comprise a single family of enzymes which evolved divergently from a primordial murein amidase.

Structural differences between the murein peptidases may reflect the structures of their corresponding peptide substrates and the origins of the different classes of serine ß-lactamases. The position that corresponds to the α-carbon on the 6-ß-aminoacyl side-chain of penicillin is not substituted in the LD-peptide (figure 2-2). Generally, ß-lactams that lack substituents at this position exhibit higher turnover numbers than their substituted counterparts due to hydrolysis by class C ß-lactamases [26]. The major secondary structural elements of DD-peptidases and serine ß-lactamases can be superposed on each other. An important structural difference, which may somehow influence substrate recognition, is the inverted orientation of helix α10 in the class C ß-lactamase [5]. These relationships support the concept that the AmpC protein may recognize both ß-lactams and LD-peptides as substrates.

An alternate LD-peptide hypothesis was originally proposed by Wise and Park; they demonstrated that penicillin is a structural analogue of the L-ala-D-γ-glu peptide [27]. However, L-ala-D-γ-glu is not a substrate for any known peptidases and it is unable to adopt the conformation that mimics penicillin and D-ala-D-ala [28]. The fact that the LD-carboxypeptidase is not inactivated by penicillin [6] may either reflect a conformational difference between penicillin and (L)-*m*-A_2pm-D-ala or an ability of the LD-carboxypeptidase to hydrolyze penicillin. Pollock originally proposed that some ß-lactamases may have retained an additional metabolic function which has escaped detection [29]. The overriding importance of the AmpC protein as a ß-lactamase may reflect a relatively benign function of (L)-*m*-A_2pm-(D)-*m*-A_2pm peptides in cell-wall metabolism. Additionally, Saz suggested a role for ß-lactamases in cell-wall turnover based on studies of an inducible chromosomal ß-lactamase in sporeforming bacteria [30].

In many bacteria related to *E. coli*, the chromosomal *ampC* gene is inducible and relationships between *ampC* induction and murein metabolism are evident.

5. CONTROL OF THE INDUCIBLE AmpC ß-LACTAMASE

In Gram-negative bacteria such as *Enterobacter cloacae*, *Citrobacter freundii* and *Pseudomonas aeruginosa*, the chromosomal *ampC* gene is induced by ß-lactam antibiotics [31,32,33]. Induction depends on the *ampR* gene which separates *ampC* from *frd* and thereby replaces the *ampA* attenuator of *E. coli*. The *ampRC* locus constitutes a divergently transcribed control unit wherein the AmpR protein binds specifically to the intergenic operator and represses transcription of both genes [31]. In the presence of ß-lactam compounds, AmpR is somehow converted to an activator of *ampC* gene expression. AmpR belongs to the LysR family of DNA binding transcriptional regulators, many of which are controlled by binding specific ligands [34]. It is unlikely that penicillin controls AmpR directly because the primary structure of AmpR lacks the ß-lactam binding motifs that are common among the PBPs and the serine ß-lactamases [31]. Additionally, induction of the *ampC* gene does not require that ß-lactams penetrate the cytoplasmic membrane [35]. Therefore, a signal transducing second messenger probably mediates information from the cell envelope to the *ampC* transcriptional apparatus.

Mutations that affect control of the inducible *ampC* gene have revealed important aspects of the induction mechanism. The *ampD* mutation has been identified both in *E. coli* transformants that harbor the cloned *ampRC* locus in a recombinant plasmid and in *Enterobacter cloacae* [36,37]. The *ampRC* transformants of *E. coli* normally express the plasmid encoded *ampC* gene in a ß-lactam inducible manner. In the presence of ß-lactam antibiotics, the *ampD* mutation is selected at a high frequency in the inducible bacteria; these mutants exhibit stably derepressed AmpC expression and extreme resistance to ß-

lactams [36,37]. In *E. coli*, the *ampD* mutation maps to an operon that encodes an additional downstream gene named *ampE* [38,39]. The AmpE amino acid sequence shares similarities with a family of membrane localized bacterial transport proteins, which include the oligopeptide permease (Opp), whereas AmpD is a cytoplasmic protein [39].

The non-inducible *ampC* gene of the *E. coli* chromosome is unaffected by the *ampDE* operon. However, *ampD* mutants of *E. coli* accumulate cell-wall peptides in the growth medium [40]. Additionally, in the simultaneous presence of exogenous A_2pm and a functional *ampD* gene, *E. coli* increases the number of pentapeptides in the murein; this implicates *ampD* in the general regulation of murein carboxypeptidase activity [40]. In the inducible bacteria, A_2pm functions as a non-specific inducer of AmpC expression [41]. These data suggest a dual role for the *ampDE* operon in control of cell-wall peptide metabolism and regulation of AmpC induction.

6. INDIRECT INDUCTION OF AmpC BY PEPTIDOGLYCAN

In the final stages of peptidoglycan metabolism, *E.coli* liberates cell-wall peptides from the glycan network of the murein due to the action of MurNAc-L-ala amidase [42]. Although some of the free cell-wall peptides diffuse through the outer membrane porins, most are recycled by the bacterium. Growing cells process the free cell-wall peptides to L-ala-D-γ-glu-L-*m*-A_2pm (tripeptide) which is recovered in the cytoplasm as a uridine diphosphate (UDP) activated peptidoglycan precursor (UDP-MurNAc-tripeptide) [43]. Therefore, cell-wall peptides are recycled by a pathway that requires a membrane bound tripeptide permease and a cytoplasmic UDP-MurNAc:tripeptide ligase.

Is it possible that the *ampDE* operon specifies enzymes of cell-wall peptide recycling? The accumulation of cell-wall peptides found in *ampD* mutants can be explained if the AmpD protein functions as a UDP-MurNAc:tripeptide ligase; by

conveying tripeptides to metabolism, AmpD could determine the tripeptide concentration. This postulate fits into a model for control of the inducible *ampC* gene. If the AmpR protein is regulated by binding the tripeptide, or a tripeptide metabolite, the ligand concentration would fluctuate in response to events that alter murein metabolism. Conceivably, ß-lactams could indirectly induce the AmpC protein by disrupting murein synthesis and stimulating the murein hydrolases which produce the tripeptide. Therefore, a product of murein hydrolase activity is modeled into a regulatory circuit which controls AmpC synthesis (figure 2-3).

Although the Opp protein contributes to the uptake of cell-wall peptides, the primary role of Opp is in the transport of protein derived peptides with L-amino acids and α-peptide bonds [44]. The unusual structure of cell-wall peptides may necessitate a unique permease under conditions where Opp activity is saturated. A role for AmpE as a permease with specificity for tripeptides can explain conflicting reports that AmpE is either essential for AmpC induction [38] or dispensable [39]. In this way, the *ampDE* operon could exert dual control over cell-wall peptide recycling and induction of the AmpC protein.

AmpC induction depends on several additional genes which influence the structure of the cell envelope. At the time of cell-division in *E. coli*, the *pbpA* and *ftsZ* gene products are required to divert peptidoglycan synthesis from the lateral walls to the septum; temperature sensitive mutants in each of these genes fail to induce the AmpC protein from a cloned *ampRC* plasmid at the non-permissive temperature [45,46]. The *ampG* mutation, unlinked from *ampC* and *ampDE*, exhibits a stably repressed phenotype both in *Enterobacter cloacae* and the *ampRC* transformants of *E. coli* [47,48]. The AmpG protein carries several putative *trans* membrane regions and is therefore most likely an integral cytoplasmic membrane protein possibly acting as a ligand transporter (Staffan Normark, personal communication). The AmpG protein may also interact directly with AmpR [48]. Whether the AmpR protein controls *ampC* induction by

sensing cell-wall peptides or other components of murein metabolism remains to be discerned. However, these results lend support to the concept that AmpC induction may occur indirectly via peptidoglycan metabolites.

7. UNIFYING PERSPECTIVE

The AmpR protein and the *ampA* attenuator exert unique mechanisms of control over *ampC* gene expression; how can they be rationalized in terms of the same AmpC gene product? An important relationship between the two control processes is that the Stringent Response regulates the biogenesis of ribosomes, the synthesis of peptidoglycan, and the murein-lipoprotein [49]. Additionally, the product of the *lov* gene helps coordinate the translational apparatus with the peptidoglycan synthetic machinery [50]. Therefore, control by growth rate could be exerted by mechanisms that monitor either the ribosomes or the peptidoglycan. By sensing peptidoglycan, AmpR could accommodate the AmpC protein both as an LD-peptidase and a ß-lactamase through a dual response to growth rate and ß-lactam inducers. Since the ribosomes are not directly sensitive to penicillin, the genetically inexpensive *ampA* attenuator may have evolved to employ the AmpC protein solely as an LD-peptidase in the absence of any selective challenge by ß-lactam compounds. In this way, *ampC* gene expression is rationalized in terms of control processes which exploit a functional duality in the AmpC protein. Indeed, a growing body of evidence demonstrates coordinate control of murein structure and AmpC synthesis. We hope the LD-peptidase analogy will be of value both in discerning the physiological function of the AmpC protein and in the rational design of improved chemotherapeutic agents. In the treatment of many infectious diseases, effective inhibitors of AmpC expression and activity are urgently required.

8. RETROSPECTIVE

The preceding hypothesis was published in 1992 and provided a foundation for the experimental work presented in the following chapters. The findings both supported and refuted certain aspects of the hypothesis, and so it is helpful to clarify these points before proceeding further. Firstly, the concept that AmpR is regulated by a recyclable cell-wall peptide was firmly established in the laboratory of Staffan Normark by *in vitro* transcription studies, which utilized an AmpR purification procedure that I developed and describe in Chapter 3. The tripeptide recycling pathway ultimately proved to be rudimentary, and a more complex pathway involving 1,6-anhydro-muropeptides was later shown to be responsible for AmpR regulation; this is described in detail in Chapter 5. Secondly, a D-stereospecific endopeptidase with ß-lactamase activity was isolated from *Bacillus cereus* and shown to be homologous to class-C ß-lactamases, which provided support for the endopeptidase hypothesis. However, my results demonstrate that the enterobacterial AmpC protein is NOT an LD-endopeptidase. Chapter 4 presents physiological findings that were interpreted to result either from an LD-endopeptidase activity of AmpC or from a non-specific effect of *ampC* gene expression. The results in Chapter 5 refute the LD-endopeptidase interpretation and uncover an unusual phenotype due to an uncharacterized genomic region, which became the subject of investigation for later chapters. Therefore, I hope the reader will not be surprised to learn that the following results culminate in a revised model and a new thesis problem.

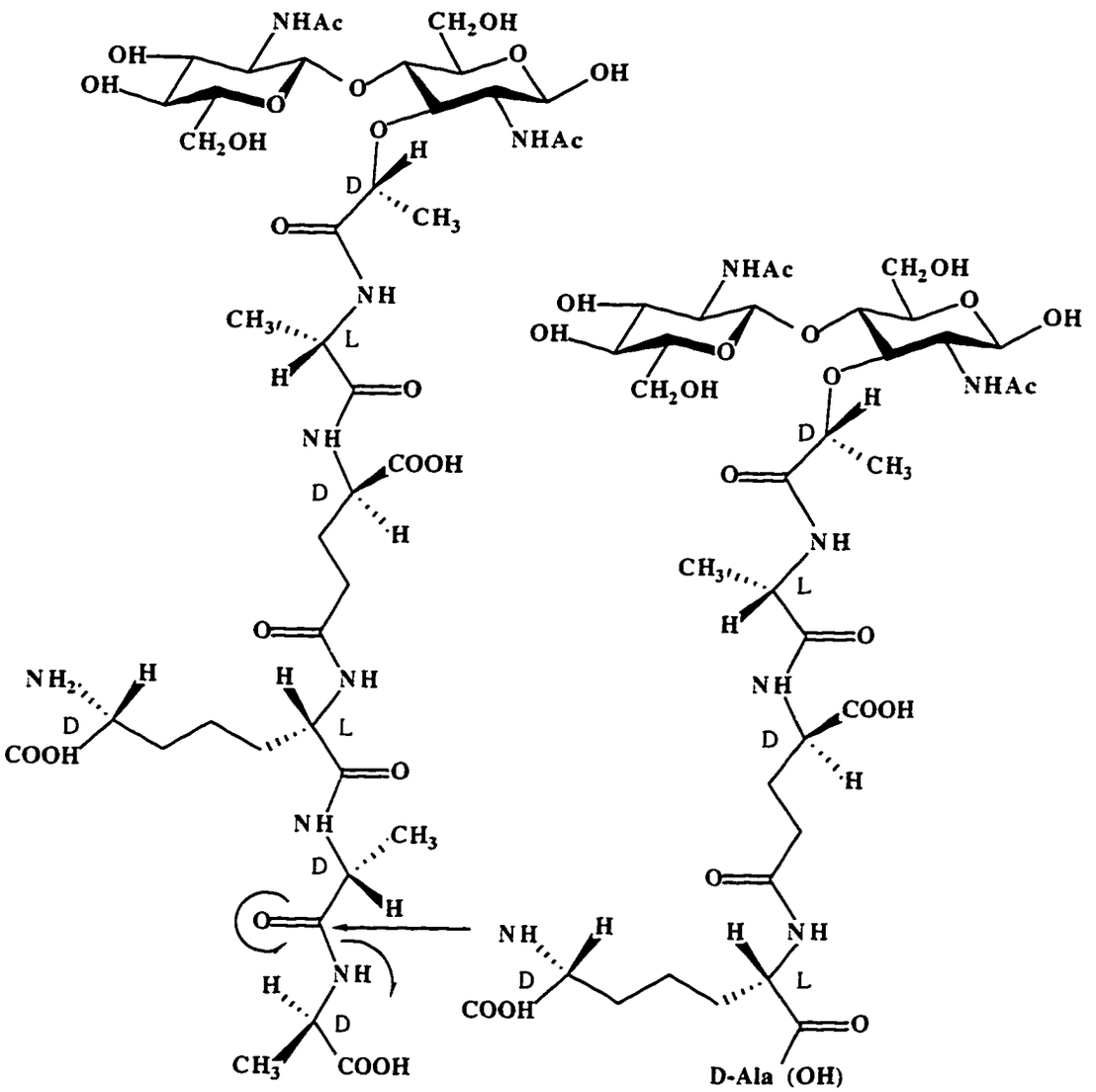


Figure 2-1:

Structure of the principal disaccharide-pentapeptide precursor that is incorporated into the murein layer of *Escherichia coli.* DD-transpeptidases catalyze acyl-transfer between the carbonyl carbon of the C-terminal D-ala-D-ala peptide and the ε-amino group of a neighboring *m*-A_2pm residue to form a peptide cross-link. DD-carboxypeptidases transfer the acyl group to water and a DD-endopeptidase ruptures the peptide cross-link (see text for discussion).

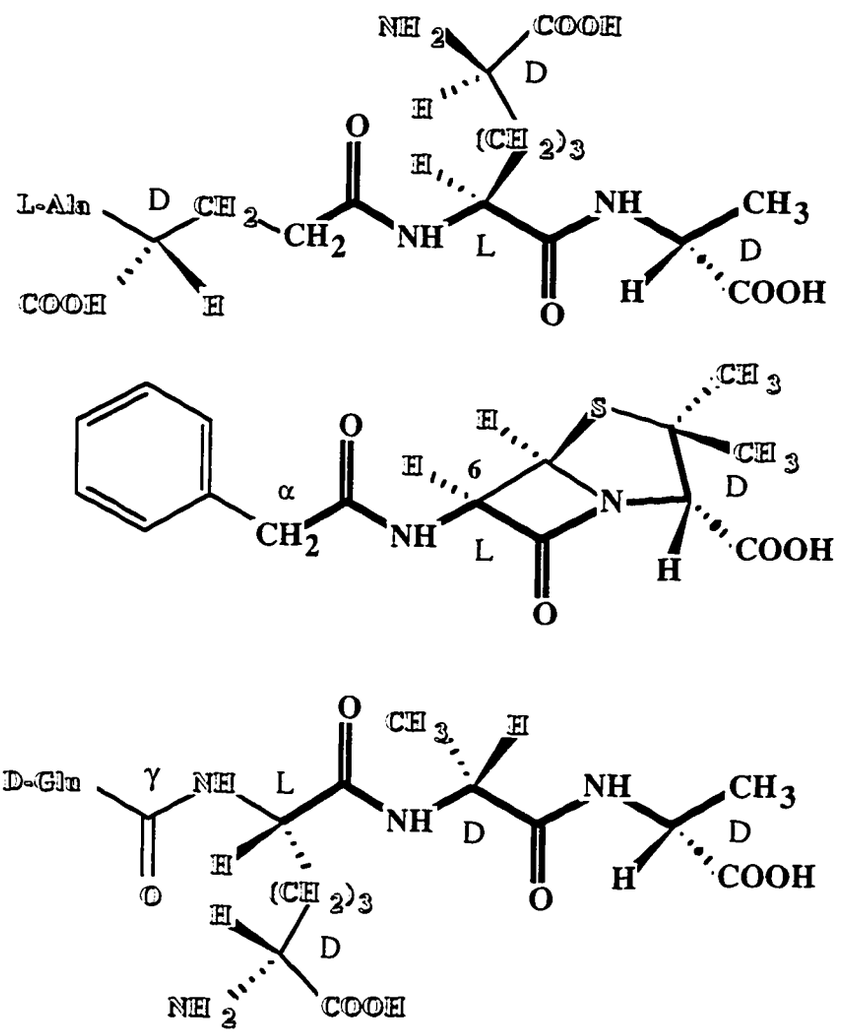


Figure 2-2:
Structural similarity between cell-wall peptides and penicillin. Enboldened atoms in the α-carbon backbones of the L,D-peptide (top) and the D,D-peptide (bottom) can be superposed on the reactive amide bond in the ß-lactam ring of benzyl-penicillin (middle). Substituents that differ between the three amide classes are outlined (see text for discussion).

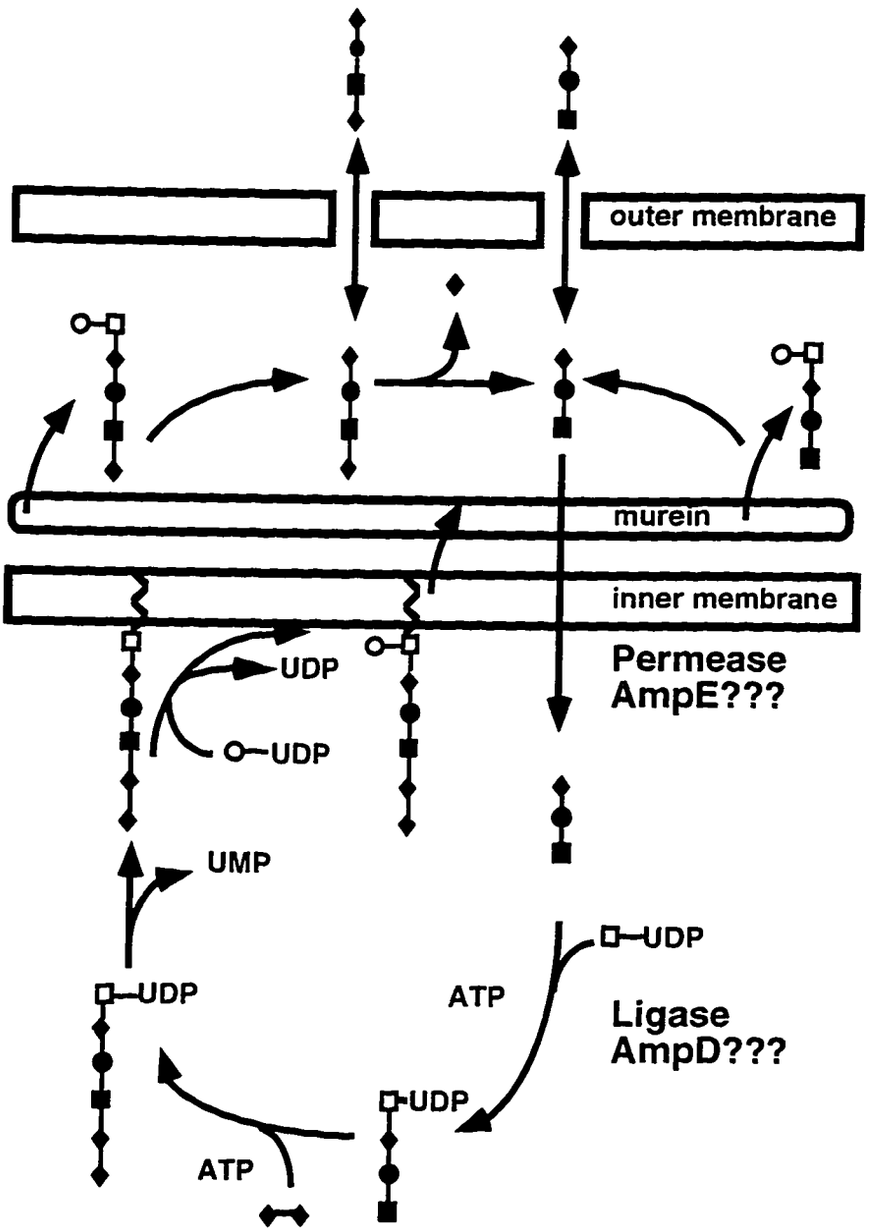


Figure 2-3:

Recycling of cell-wall peptides in *E. coli*. Peptides are liberated from the murein and processed to the tripeptide L-ala-D-γ-glu-(L)-m-A_2pm (tripeptide) which is coupled to a nucleotide activated precursor of peptidoglycan biosynthesis. The process requires a tripeptide permease and a UDP-MurNAc:tripeptide ligase. Symbols: ○ GlcNAc; □MurNAc; ♦ L-Ala; ● D-Glu; ■*meso*-A_2pm.

(See text for discussion. Modified from [43] with permission)

References:

[1] Park, J.T. (1987) in: *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (Neidhardt, C. ed) pp. 23-30 and 663-671, American Society for Microbiology, Washington, D.C.

[2] Höltje, J-.V., and Tuomanen, E.I. (1991) J. Gen. Microbiol. 137, 441-454.

[3] Tipper, D.J., and Strominger, J.L. (1965) Proc. Natl. Acad. Sci. USA. 54, 1133-1141.

[4] Ghuysen, J-.M. (1991) Ann. Rev. Microbiol. 45, 37-67.

[5] Joris, B., Ledent, P., Dideberg, O., Fonzé, E., Lamotte-Brasseur, J., Kelly, J., Ghuysen, J-.M., and Frère, J.M. (1991) Antimicrob. Agents Chemother. 35, 2294-2301.

[6] Ursinus, A., Steinhaus, H., and Höltje, J-.V. (1992) J. Bacteriol. 174, 441-446.

[7] Metz, R.S., Henning, S., and Hammes, W.P. (1986) Arch. Microbiol. 144, 175-180.

[8] Begg, K.J., Takasuga, A., Edwards, D.H., Dewar, S.J., Spratt, B.G., Adachi, H., Takahisa, O., Hiroshi, M., and Donachie, W.D. (1990) J. Bacteriol. 172, 6697-6703.

[9] Beck, B.D., and Park, J.T. (1976) J. Bacteriol. 126, 1250-1260.

[10] Braun, V. (1975) Biochim. Biophys. Acta 415, 335-377.

[11] Torti, S.V., and Park, J.T. (1976) Nature 263, 323-326.

[12] Weigand, R.A., Vinci, K.D., and Rothfield, L.I. (1976) Proc. Natl. Acad. Sci. USA. 73, 1882-1886.

[13] Hirota, Y., Suzuki, H., Nishimura, Y., and Yasuda, S. (1977) Proc. Natl. Acad. Sci. USA. 74, 1417-1420.

[14] Höltje, J.-V., and Glauner, B. (1990) Res. Microbiol. 141, 75-89.

[15] Glauner, B., Höltje, J-.V., and Schwarz, U. (1988) J. Biol. Chem. 263, 10088-10095.

[16] Glauner, B., and Höltje, J.-V. (1990) J. Biol. Chem. 265, 18988-18996.

[17] Driehuis, F., and Wouters, J.T.M. (1987) J. Bacteriol. 169, 97-101.

[18] Jaurin, B., and Normark, S. (1979) J. Bacteriol. 138, 896-902.

[19] Grundström, T., and Jaurin, B. (1981) Proc. Natl. Acad. Sci. USA. 79, 1111-1115.

[20] Jaurin, B., Grundström, T., Edlund, T., and Normark, S. (1981) Nature 290, 221-225.

[21] Normark, S., Edlund, T., Grundström, T., Bergström, S., and Wolf-Watz, H. (1977) J. Bacteriol. 132, 912-922.

[22] Olsson, O., Bergström, S., and Normark, S. (1982) EMBO J. 1, 1411-1416.

[23] Tuomanen, E., Cozens, R. Tosch, W., Zak, O., and Tomasz, A. (1986) J. Gen. Microbiol. 132, 1297-1304.

[24] Trueba, F.J., and Woldringh, C.L. (1980) J. Bacteriol. 142, 869-878.

[25] Aldea, M., Herrero, E., and Trueba, F.J. (1982) Current Microbiol. 7, 165-168.

[26] Frère, J.M., Galleni, M., Joris, B., DeMeester, F., Charlier, P., and Dideberg, O. (1989) in: Recent Advances in the Chemistry of ß-Lactam Antibiotics (Bentley, P.H., and Southgate, R. eds) pp. 321-332, Royal Society of Chemistry, London.

[27] Wise, E.M., and Park, J.T. (1965) Proc. Natl. Acad. Sci. USA. 54, 75-81.

[28] Virudachalam, R., and Rao, V.S.R. (1977) Int. J. Peptide Protein Res. 10, 51-59.

[29] Pollock, M.R. (1967) Brit. Med. J. 4, 71-77.

[30] Saz, A.K. (1970) J. Cell. Physiol. 76, 397-404.

[31] Lindquist, S., Lindberg, F., and Normark, S. (1989) J. Bacteriol. 171, 3746-3753.

[32] Honoré, N., Nicolas, M-.H., and Cole, S.T. (1986) EMBO J. 5, 3709-3714.

[33] Lodge, J.M., Minchin, S.D., Piddock, L.J.V., and Busby, J.W. (1990) Biochem. J. 272, 627-631.

[34] Henikoff, S., Haughn, G.W., Calvo, J.M., and Wallace, J.C. (1986) Proc. Natl. Acad. Sci. USA. 85, 6602-6606.

[35] Everett, M.J., Chopra, I., and Bennett, P.M. (1990) Antimicrob. Agents Chemother. 34, 2429-2430.

[36] Lindberg, F., Lindquist, S., and Normark, S. (1987) J. Bacteriol. 169, 1923-1928.

[37] Nicolas, M-.H., Honoré, N., Jarlier, V., Phillipon, A., and Cole, S.T. (1987) Antimicrob. Agents Chemother. 31, 1091-1102.

[38] Honoré, N., Nicolas, M-.H., and Cole, S.T. (1989) Molec. Microbiol. 3, 1121-1130.

[39] Lindquist, S., Galleni, M., Lindberg, F., and Normark, S. (1989) Molec. Microbiol. 3, 1091-1102.

[40] Tuomanen, E., Lindquist, S., Sande, S., Galleni, M., Light, K., Gage, D., and Normark, S. (1991) Science 251, 201-204.

[41] Sanders, C.C. (1990) Program Abstr. 30th Intersci. Conf. Antimicrob. Agents Chemother., abstr. 274, pp. 132.

[42] van Heijenoort, J., Parquet, C., Flouret, B., and van Heijenoort, Y. (1975) Eur. J. Biochem. 58, 611-619.

[43] Goodell, E.W. (1985) J. Bacteriol. 163, 305-310.

[44] Goodell, E.W., and Higgins, C.F. (1987) J. Bacteriol. 169, 3861-3865.

[45] Oliva, B., Bennett, P.M., and Chopra, I. (1989) Antimicrob. Agents Chemother. 33, 1116-1117.

[46] Ottolenghi, A.C., and Ayala, J.A. (1991) Antimicrob. Agents Chemother. 35, 2359-2365.

[47] Korfmann, G., and Sanders, C.C. (1989) Antimicrob. Agents Chemother. 33, 1946-1951.

[48] Bartowsky, E., and Normark, S. (1991) Molec. Microbiol. 5, 1715-1725.

[49] Vanderwel, D., and Ishiguro, E.E. (1984) Can. J. Microbiol. 30. 1239-1246.

[50] Bouloc, P., Jaffé, A., and D'Ari, R. (1989) EMBO J. 8, 317-323.

CHAPTER 3

Overproduction, Solubilization, Purification and DNA-Binding Properties of AmpR from *Citrobacter freundii*[1]

When one publishes an hypothesis, one is sentenced to hard labor.

André Lwoff

[1]A version of this chapter has been published. Bishop, R.E., and Weiner, J.H. (1993) *Eur. J. Biochem.* **213:** 405

INTRODUCTION

Escherichia coli harbors a class C ß-lactamase (AmpC) encoded by the *ampC* gene which overlaps the fumarate reductase operon (*frd*) at 94.5' on the *E. coli* chromosome (Grundström and Jaurin, 1982). The *ampC* promoter resides within the *frd* coding sequence and directs transcription from upstream of the *frd* terminator which acts as a growth rate dependent attenuator of *ampC* transcription (Jaurin et al., 1981). Hence, the *ampC* gene is expressed poorly and is not normally an agent of ß-lactam resistance. The relationship between growth rate and AmpC function is not understood, but it has been demonstrated that *ampC* gene expression appears to be coordinated with cell-wall LD-peptidase metabolism in *E. coli* (Bishop and Weiner, 1992).

Other members of the *enterobacteriaceae*, including *Citrobacter freundii* (Lindquist et al., 1989) and *Enterobacter cloacae* (Honoré et al., 1986), harbor an *ampC* gene which is separated from *frd* by a divergently transcribed regulatory gene (*ampR*). The AmpR protein belongs to the LysR family of prokaryotic DNA-binding transcriptional regulators (Henikoff et al., 1988) and renders the *ampC* gene inducible by ß-lactam compounds. In the absence of ß-lactam inducers, AmpR binds an operator DNA sequence located in the *ampRC* intergenic region and represses transcription of both genes (Lindquist et al., 1989). In the presence of ß-lactams, AmpR is converted to an activator of *ampC* transcription. However, the mechanism by which AmpR senses ß-lactams is unknown.

LysR members share significant amino acid sequence similarity in their N-terminal domains which possess a putative DNA-binding helix-turn-helix (HTH) amino acid sequence motif; sequence divergence at the C-terminus may reflect an inducer binding domain since several LysR members are known to activate transcription only in the presence of a specific low molecular weight ligand (Henikoff et al., 1988). In the case of AmpR, it is not necessary that ß-lactams enter the cytoplasm to induce *ampC* transcription (Everett et al., 1990) nor does AmpR exhibit any sequence motifs of the

known penicillin-binding proteins. Therefore, it is probable that AmpR interacts with a second messenger molecule produced in response to ß-lactam antibiotics.

Unlinked mutations that deregulate the *ampRC* control unit may reflect key elements of the signal transduction pathway; these loci include *ampDE* (Honoré et al., 1989), *ampG* (Korfmann and Sanders, 1989), *pbpA* (Oliva et al., 1989), and *ftsZ* (Ottolenghi and Ayala, 1991). The AmpD, PbpA, and FtsZ gene products control aspects of cell-wall metabolism, whereas the AmpE and AmpG proteins share amino acid sequence similarities with various membrane transport proteins (reviewed in Bishop and Weiner, 1992). Additionally, mutations in the *ampR* gene can suppress the *ampG* mutation which suggests that the AmpR and AmpG proteins may interact physically (Bartowsky and Normark, 1991). Therefore, it appears that AmpR may be sensitive to components of the cell-envelope which are disrupted by the action of ß-lactam antibiotics.

In order to characterize the signaling mechanism of *ampC* induction, it is necessary to understand the biochemical properties of the AmpR protein and its mode of sequence-specific DNA recognition. Although reports of the expression and purification of the *C. freundii* AmpR protein (CfAmpR) have appeared (Bishop et al., 1991; Bartowsky and Normark, 1991), the physical properties of the purified protein have not been characterized. In this report, we present an improved procedure for the expression and purification of CfAmpR and describe the solubility and DNA binding properties of the purified protein.

EXPERIMENTAL PROCEDURES

Bacteria and Plasmids

Escherichia coli strain HB101 and plasmid pBR322 (Sambrook et al., 1989) were used for cloning procedures; strain JRG582 (Langley and Guest, 1977) was obtained from Dr. S.T. Cole as was plasmid pEc1c (Nicolas et al., 1987). Plasmid pNU302 was

obtained from Dr. S. Normark (Lindberg et al., 1985), and expression vector pJF118EH (Fürste et al., 1986) was provided by Dr. B.D. Lemire. Strains and plasmids are described in Table 3-1.

Materials

Restriction enzymes, DNA polymerase from *Thermus aquaticus* (Taq), and bacteriophage T4 kinase were purchased from BRL, Inc. Oligonucleotides were synthesized by Dr. K. Roy in the University of Alberta Oligonucleotide Core Facility on an Applied Biosystems 391 automated synthesizer. Radionuclides were from Dupont and all other chemicals were from commercial sources.

Media and growth conditions

E. coli strains were grown on Luria-Bertani media (Miller, 1972) supplemented with 10 μg/mL thiamine at 37°C. Bacteria harboring plasmids were selected on media (liquid or plates) containing 100 μg/mL ampicillin.

Recombinant DNA procedures

Restriction enzyme digestions, ligations, transformations, mini plasmid preparations, and electrophoresis were carried out as described (Sambrook et al., 1989).

Protein determination

Protein concentrations were determined by the method of Lowry (Lowry et al., 1951) for routine analysis. Fractions eluted from columns were monitored by following the absorbance at A_{260} and/or A_{280} (Stoschek, 1990).

Polymerase chain reaction mutagenesis

The polymerase chain reaction (PCR) was performed in a 100 μL reaction volume containing 10 ng template DNA, 1 pM each of the 5' and 3' oligonucleotide primers (see Figure 1), 0.2 mM each of deoxynucleotide triphosphates (dATP, dGTP, dCTP, dTTP; prepared as described by Sambrook et al., 1989), and 0.1 vol of 10X PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 0.01% gelatin). The reaction was initiated with 2.5 U of Taq polymerase, overlayed with 100 μL of mineral oil, and subject

to 25 cycles through 30 s at 94 °C, 30 s at 72 °C, and 60 s at the lowest of the two calculated annealing temperatures for oligonucleotide primers as determined from sequences which are complementary to the target DNA (Mullis and Faloona, 1987). The annealing temperatures for the complementary sequences were determined empirically with the equation, temperature (°C)=4(G+C)+2(A+T). Thermocycling was performed with a Techne model PHC-2 Dri-Block thermocycler.

Polyacrylamide gel electrophoresis

Protein samples were solubilized in an equal volume of 2X Laemmli SDS buffer according to Laemmli (1970) without boiling. SDS-PAGE on 1.5 mm thick 12% acrylamide slab gels was performed with a Bio-Rad Mini-PROTEAN II Dual Slab Cell prepared according to the manufacturers instructions and operated at a constant 180 Volts for 45 min. Gels were stained with Coomassie Blue dye and destained with 10% acetic acid. Molecular mass standards include α-lactalbumin (14.2kDa), trypsin inhibitor (20.1kDa), carbonic anhydrase (29kDa), and ovalbumin (45kDa).

When necessary, protein samples contained in high salt buffers were precipitated with trichloroacetic acid prior to solubilization. Briefly, protein extracts were diluted into a 0.5 mL solution containing 0.03% sodium deoxycholate and 5% trichloroacetic acid and cooled on ice for 10 min. Samples were centrifuged at 12,000Xg for 5 min, drained, and the precipitate resuspended in 25 μL 0.5N NaOH. Samples were then solubilized with an equal volume of 2X Laemmli solubilization buffer and subject to SDS-PAGE as described above. As the deoxycholate affects protein migration slightly, all samples including the molecular mass standards were subject to trichloroacetic acid precipitation.

Protein expression and solubilization

Starter cultures of JRG582/pJF118EH, JRG582/pJFCfAmpR and JRG582/pJFEcAmpR were grown overnight to stationary phase in Luria-Bertani medium; a 1% inoculum was then subcultured on a 1L scale and grown an additional 2 h to mid-logarithmic phase (A_{600}=0.5). The culture was made 1 mM in isopropyl-ß-D-

thiogalactopyranoside (IPTG) and grown an additional 1 h before cells were harvested, washed with 10 mL 50 mM MOPS-KOH pH7, and frozen in liquid nitrogen. Subsequent steps were performed at 4°C.

The cells were thawed and suspended in 10 mL 50 mM MOPS-KOH pH7 and passed twice through a French Pressure Cell at 16000 psi (110 MPa). Extracts containing AmpR were centrifuged (30,000Xg for 10min) in a Beckman TL-100 ultracentrifuge (Ti 100.3 rotor) and the pellet washed with 4 mL 20 mM Tris-HCl pH8 containing 5 mM EDTA and 1 mM dithiothreitol. The pellet was resuspended in 2 mL of Buffer A-500 (500 mM NaCl, 10 mM Tris-borate pH8, 10% glycerol, 1 mM dithiothreitol), transferred to a 15 mL conical tube and sonicated twice for 15 s at 100 Watts in a Braun-sonic model 1510 sonicator. To the sonicate was added 8 mL of Buffer B (1.65 M guanidine-HCl [Gdn/HCl], 10 mM Tris-borate pH8, 10% glycerol, 1 mM dithiothreitol), the suspension was incubated on ice for 30 min, centrifuged as above and the supernatant isolated for further analysis.

Protein purification

The solubilized AmpR protein of *Citrobacter freundii* was dialyzed exhaustively against Buffer A-500, centrifuged, and the supernatant (approximately 10 mL at 0.5-0.8mg/mL protein) was further dialyzed against 4L Buffer A-50 (50 mM NaCl, 10 mM Tris-borate pH8, 10% glycerol, 1 mM dithiothreitol) until a white precipitate formed (2-5 h). The dialysate was centrifuged and the white precipitate washed with 4 mL of Buffer A-50, resuspended in 2 mL of Buffer A-500, resolubilized on addition of 8 mL of Buffer B, and dialysed exhaustively against Buffer A-1000 (1 M NaCl, 10 mM Tris-borate pH8, 10% glycerol, 1 mM dithiothreitol). The dialysate was centrifuged and the supernatant (approximately 10 mL at 0.3-0.5 mg/mL protein) was loaded onto a gel-filtration column.

Routine gel-filtration was performed at 4°C using a 1L Sephacryl S-200 HR gel-filtration column (150X3 cm) pre-equilibrated with Buffer A-1000 and run at a flow rate of 0.5 mL/min. Fractions containing protein or nucleic acid were identified by the

A_{260}/A_{280} ratio (Warburg and Christian, 1942). The column was typically loaded with protein at 1% of the column volume and the peak protein fractions were pooled (approximately 15 mL at 0.15-0.3 mg/mL protein). The purified CfAmpR protein was dialyzed exhaustively against a DNA binding assay buffer, Buffer C (300 mM potassium glutamate, 10 mM Tris pH8.5, 30% glycerol, 1 mM ß-mercaptoethanol), and then centrifuged; the supernatant was somewhat concentrated (approximately 8 mL at 0.25-0.5 mg/mL protein). Therefore, several mg of CfAmpR was purified from a 1L culture of induced bacteria.

Analytical gel-filtration was performed on a Pharmacia FPLC system with a Superose 12 prepacked HR 10/30 column (24 mL volume). The FPLC system was operated under ambient conditions at a flow rate of 0.5 mL/min and monitored by following the absorbance of the eluate at 280 nm; the absorbance readings were stabilized when the column was pre-equilibrated and eluted with Buffer A-1000 in which 1 mM ß-mercaptoethanol was substituted for dithiothreitol. Molecular mass standards included bovine serum albumin (67kDa), ovalbumin (43kDa), chymotrypsin (25kDa), and ribonuclease (13.7kDa).

AmpR DNA-binding assay

A 1.5 kb *Bam*H1 fragment containing the *C. freundii ampR* operator sequence from plasmid pNU302 was ligated into the single *Bam*H1 site of plasmid pBR322 to generate plasmid pREB11. Plasmid pREB11 was digested with *Bam*H1, subjected to phenol-chloroform extraction and ethanol precipitated as described by Sambrook et al. (1989). The DNA was resuspended in water at a concentration of 100 ng/µL for use in non-radioactive DNA-binding assays. 100 ng of the DNA was added to 10 µL of Buffer C containing various dilutions of purified AmpR protein. Samples were allowed to incubate for 30 min at 25°C and subjected to electrophoresis on a 5% polyacrylamide gel (30:0.8, acrylamide-bisacrylamide) in TBE buffer (89 mM Tris-borate pH8, 2 mM

EDTA). Gels were stained with ethidium bromide and illuminated under ultraviolet light (Sambrook et al. 1989).

Radiolabeling of oligonucleotides was performed in 20μL of T4 kinase buffer (50mM Tris-HCl pH 8, 10mM $MgCl_2$, 5 mM dithiothreitol, 0.1mM spermidine, 0.1mM EDTA) containing 20μCi γ-^{32}P-ATP (10 μCi/nmol), 10U T4 kinase, 25ng of annealed DNA and incubated at 37°C for 30 min. The AmpR DNA-binding assay was performed as described above, except that 1.25 ng of radiolabeled duplex oligonucleotide was incubated in the presence of 100ng of dTC/dGA non-specific DNA and 9% polyacrylamide gels were exposed by autoradiography.

RESULTS

PCR mutagenesis and ampR gene expression

AmpR represses its own transcription by binding to an operator sequence which overlaps the *ampR* promoter (Lindquist et al., 1989). In our initial attempts to replace the *ampR* promoter with an IPTG inducible *tac*-promoter, we failed to rapidly overproduce AmpR protein when it was expressed with the wild-type ribosome binding sequence (rbs). The rbs between the promoter and translational start signals of the *ampR* genes from *Enterobacter cloacae* and *Citrobacter freundii* do not possess significant complementarity with the 16S rRNA and are separated from the initiating Met codon by a sub-optimal spacing of 5 bp. This prompted us to enhance the efficiency of translation by incorporating an efficient rbs (rbs*) into the *ampR* genes; the rbs* was designed according to the criteria of Gold and Stormo (1987).

The PCR mutagenesis approach was adopted in order to amplify the *ampR* gene sequences with modified termini containing restriction enzyme sites, to facilitate cloning, and 5' ends with the rbs* sequence (Figure 3-1). The amplified *ampR* constructs were cloned behind the IPTG inducible *tac*-promoter of expression vector pJF118EH. *E. coli* strain JRG582 was transformed with the cloned PCR constructs from *Citrobacter freundii*

(pJFCfAmpR) or *Enterobacter cloacae* (pJFEcAmpR) as well as pJF118EH for use as a control. Cells grown to mid-logarithmic phase were induced with IPTG and protein extracts prepared and analyzed by SDS-PAGE as described in Experimental Procedures (Figure 3-2); after 1h, the AmpR expressing cells produced a unique band near 32kD which corresponds to the predicted molecular mass of the AmpR protein from each of *C. freundii* and *E. cloacae*. The intensity of the AmpR band reached its peak after 1h and accounts for up to 10% of total cellular protein as estimated from electrophoresis (Figure 3-2, lanes B and C)

Solubilization of AmpR aggregates

Lysates from AmpR expressing cells were found to contain the bulk of the AmpR protein as an insoluble aggregate which could be collected by centrifugation at 30, 000Xg (Figure 3-3). The aggregates were enriched in AmpR protein and accounted for approximately 30% of insoluble protein as estimated from electrophoresis (Figure 3-3, lanes C and c). Despite similar primary structure and physiological function, the two AmpR proteins exhibited unique solubility properties. The CfAmpR protein could be solubilized in a buffer containing a final concentration of 1.32M Gdn/HCl (see Experimental Procedures; Figure 3-3, lane D) and accounted for nearly 50% of soluble protein; the EcAmpR protein was insoluble in 1.32M Gdn/HCl (Figure 3-3, lane d).

Purification of AmpR from C. freundii

CfAmpR remained soluble when the Gdn/HCl was replaced by dialysis with Buffer A-500 (Figure 3-4, lane D), which contains 500 mM NaCl (see Experimental Procedures), and accounted for greater than 50% of the soluble protein; CfAmpR was unstable and precipitated out of solution over a period of several days at 4°C. This process could be accelerated by dialyzing the extracts against Buffer A-50, which contains 50 mM NaCl. The precipitate could then be resolubilized with 1.32M Gdn/HCl and remained soluble when dialyzed against Buffer A-1000 (1M NaCl). The protein was approximately 90% homogenous (Figure 3-4, lane E) at this stage. The EcAmpR protein

was refractive to all attempts at solubilization under non-denaturing conditions (S. Pawliuk, unpublished data).

The solubility of CfAmpR depended strongly on salt as well as 10% glycerol which was present throughout the solubilization and purification procedure. The stability of the CfAmpR protein in 1M NaCl also depends on the protein concentration; above 1 mg/mL the protein gradually precipitated at 4°C over a 24 h period whereas dilute solutions (<0.1 mg/mL) were stable for up to 1 week at 4°C. Typically, experiments yielded 0.3-0.5 mg/mL of resolubilized CfAmpR protein extract which remained soluble for at least 2 days at 4°C. Therefore, gel filtration chromatography was performed immediately as described in Experimental Procedures. A single protein peak was detected and isolated from the column which migrated with the AmpR band at 32kD in SDS-PAGE analysis (Figure 3-4, lanes F,G,H). Analytical gel-filtration chromatography (Figure 5) revealed a broad peak in the column void volume (8 mL; estimated as 1/3 column volume) which exhibited an absorbance ratio of A_{260}/A_{280}=1.88-1.91, indicative of nucleic acid (Warburg and Christian, 1942). As the reported exclusion limit of Superose 12 (Pharmacia) is $2x10^6$ Da for globular proteins, the nucleic acid probably represents an array of high molecular mass DNA fragments. The second monodisperse peak was protein (A_{260}/A_{280}=0.63-0.65) and migrated with a molecular mass near 64kD (Figure 3-5) which indicates that CfAmpR protein exists in solution as a dimer. The column purified CfAmpR protein gradually precipitated if left to incubate at 4°C; however, the purified precipitate could not be resolubilized again which may indicate that solubilization depends partly on the presence of DNA.

After CfAmpR was purified by gel filtration, the protein was dialyzed against a DNA binding assay buffer containing 300 mM potassium glutamate and 30% glycerol (Buffer C, Experimental Procedures). The stability of purified CfAmpR in Buffer C was dramatically improved and the protein remained soluble at concentrations between 0.1-0.5 mg/mL for over a year at -20°C.

DNA-binding specificity of C. freundii AmpR

Digestion of pREB11 with *Bam*H1 generates two DNA fragments of 1.5 kbp and 4.3 kbp which serve as specific and non-specific ligands for the CfAmpR protein. In Buffer C (Figure 3-6), CfAmpR specifically binds and retards the mobility of the 1.5 kbp DNA fragment containing the *ampR* operator sequence. Approximately half of the 1.5 kbp DNA fragment containing the *ampR* operator sequence was mobilized by 50 ng of purified CfAmpR; non-specific binding to the 4.3 kbp DNA fragment is observed above 100 ng of CfAmpR. We conclude that purified CfAmpR protein is active in sequence specific DNA binding.

An A-tract in the left half of the *C. freundii ampR* operator was aligned with similar sequences from *E. cloacae, Yersinia enterocolitica*, and *Pseudomonas aeruginosa* (Figure 3-7A); a pseudo-consensus half-operator DNA with an inverted repeat sequence that surrounds the A-tract in the *C. freundii ampR* operator was selected for DNA-binding studies. Oligonucleotides corresponding to the two strands of the pseudo-consensus *ampR* half-operator duplex DNA sequence (5'CTGTTAGAAAAACTAACAG3' and 5'CTGTTAGTTTTTCTAACAG3'), or the 38bp wild-type operator (5'CCTGTTAGAAAAACTTATATCTGCTGCTAAATTTAACC3' and 5'GGTTAAATTTAGCAGCAGATATAAGTTTTTCTAACAGG3') were annealed and then labeled with ^{32}P by T4 kinase (Experimental Procedures). A duplex 22 mer containing the eukaryotic Sp1 element was also labeled for use as a control (5'ATTCGATCGGGGCGGGGCGAGC3' and 5'GCTCGCCCCGCCCCGATCGAAT3'). Figure 3-7B shows the mobility shift of the wild-type operator (OR12) and the pseudo-consensus half-operator (OR1) mediated by purified AmpR whereas the mobility of the *Sp1* DNA (Sp1) is unaffected by the AmpR protein. Figure 3-7C demonstrates that excess non-radioactive OR12 and OR1 can effectively compete for binding with either of the AmpR-operator complexes. Although OR1 does not compete as effectively as OR12,

we conclude that purified CfAmpR is capable of specifically binding to the pseudo-consensus *ampR* half-operator *in vitro*.

DISCUSSION

The LysR family of DNA-binding transcriptional regulators control diverse biological phenomena (Henikoff et al., 1988). Several properties of CfAmpR are shared with other LysR proteins which may indicate that the techniques developed here are adaptable to other systems. For example, a translational block in expression of the *oxyR* gene (an autoregulated *lysR* family member) has been relieved through incorporation of an rbs* with annealed oligonucleotides (Storz et al., 1990). We have shown that PCR can also modify translational signals to improve expression of the AmpR protein; PCR mutagenesis provided the additional advantage of replacing the autoregulated *ampR* promoter in the absence of conveniently placed restriction enzyme sites and of incorporating new restriction enzyme sites to facilitate cloning. Additionally, the solubility of the OxyR protein depends on NaCl concentrations in a manner that has been exploited to enrich OxyR by selective precipitation (Tao et al., 1991). For *in vitro* studies, it has been reported that potassium glutamate is a preferred salt for protein-DNA interactions (Leirmo et al., 1987). We have developed a system for sequence specific DNA-binding which exploits a potassium glutamate buffer and which may be useful for stabilizing other relatively insoluble LysR proteins for DNA-binding studies. The NhaR protein (another LysR member) was shown to migrate as a tetramer by gel-filtration when crude extracts were analyzed (Schell et al., 1990). Although we have shown that purified CfAmpR exists in solution as a dimer, it is possible that higher order oligomers may form *in vivo*.

The prokaryotic HTH proteins represent an important class of sequence-specific DNA binding proteins; these dimeric proteins bind dyad symmetric DNA sequences to create complexes with twofold rotational symmetry (Steitz, 1990). Each subunit of the

dimeric protein mediates sequence-specific DNA recognition by an α-helix (the reading helix) which together bind in adjacent major grooves of the DNA. It is in the major groove where the critical DNA bases for sequence-specific DNA recognition reside; the intervening DNA sequence can modulate protein-DNA interactions by accommodating protein induced distortions in the normal B-DNA conformation (Steitz, 1990). All prokaryotic HTH proteins for which the atomic structures have been determined contain a critical glycine residue in the turn region at position 9 of the 20 residue long HTH motif; the conformation of the turn corresponds to a left handed α-helix where a glycine residue must reside (Brennan and Matthews, 1990). AmpR, and other members of the LysR family, do not contain the critical glycine at position 9 of the HTH motif. However, the reading helix of AmpR appears to function normally since the recent mutant analysis of Bartowsky and Normark (1991) demonstrated that a S35F mutant of AmpR has lost sequence-specific DNA binding activity. In AmpR, Ser^{35} corresponds to residue 13 in the HTH motif, which is suitably positioned in the reading helix for the amino acid side chain to interact with DNA bases in the major groove (Brennan and Matthews, 1990). Our results that AmpR exists in solution as a dimer is also consistent with the HTH motif.

Perhaps the most difficult factor to rationalize in terms of the HTH motif is the structure of the *ampR* operator; the *C. freundii* operator is 38 base pairs long, nearly 4 full helical turns of B-form DNA, and contains no significant dyad elements. The only outstanding feature in the *ampR* operator is an A-tract centered in the left half of the sequence. The presence of only 2 hydrogen bonds between dAdT Watson-Crick base pairs, compared with 3 hydrogen bonds between dGdC pairs, imparts a higher degree of propeller twisting which optimizes base stacking interactions and narrows the minor groove (Diekman et al., 1992). Tracts of 5 or 6 contiguous A (or T) residues intrinsically deflect the trajectory of the DNA helix by as much as 20 degrees (Koo et al., 1990). When the DNA sequences of the *ampRC* intergenic region from several enterobacteria and *Pseudomonas aeruginosa* became available (Seone et al., 1992; Lodge et al., 1990),

it was clear that A-tracts (or T-tracts) corresponding to the centre of the left half of the *ampR* operator from *C. freundii* were conserved. Alignment of the sequences with the A-tract revealed a pseudo-consensus sequence with a dyad symmetry element surrounding the A-tract (figure 3-7A); the pseudo-consensus sequence also contains a T-N_{11}-A sequence element which is conserved in other LysR operators (Goethals et al., 1992). Elements of the pseudo-consensus sequence are also conserved in the right half of the operator although the A-tract is replaced by an A/T rich sequence which corresponds to the -10 region of the *ampR* promoter. Whether the A-tract mediates initial protein-operator recognition or facilitates AmpR induced DNA bending or both is unknown. However, the demonstration that purified CfAmpR binds to the pseudo-consensus *ampR* half-operator supports the HTH model of DNA-protein recognition.

Other members of the LysR family share similarities with CfAmpR in the mode of DNA binding. For example, NodD is a dimer that binds tandem operators each of which contain a conserved dyad symmetry element surrounding a T-N_{11}-A sequence with variable internal bases (Goethals et al., 1992); these authors noted similarities between the *nodD* operator and other *lysR*-type operators. NodD activates transcription in the presence of flavonoid inducers and has been reported to induce a bend in its operator (Fisher and Long, 1992). Additionally, TrpI utilizes indole-glycerol-phosphate as inducing ligand. Chang and Crawford (1989) have shown that TrpI binds to the left half of its 40 bp operator in the absence of indole-glycerol-phosphate, but binds to both halves of the operator in the presence of indole-glycerol-phosphate. Therefore, it appears that indole-glycerol-phosphate switches TrpI from repressor to activator by promoting the occupation of both halves of the operator. It is an intriguing possibility that differential DNA-bending due to single or tandem occupation of LysR operators may determine the switch between repression and activation of transcription.

The nature of the AmpR ligand remains unknown. However, the demonstration that purified dimeric CfAmpR binds to a pseudo-consensus half-operator DNA with an

inverted repeat sequence suggests that the activated wild-type *ampR* operator may be occupied by AmpR as a dimer of dimers. Therefore, the inducing ligand should be identifiable as a species that promotes tandem occupation of the wild-type operator. Experiments to identify the nature of the AmpR ligand are currently in progress.

Table 3-1. *Escherichia coli strains and plasmids used in this study*

Strain / Plasmid	Genotype / Description	Reference
HB101	*supE*44 *hsdS*20 *recA*13 *ara*-14 *proA*2 *lacY*1 *galK*2 *rpsL*20 *xyl*-5 *mtl*-1 Highly transformable; used for general cloning procedures.	Sambrook et al. (1989)
JRG582	HfrH(Δ*aroP-nadC*) *ampDE*$^-$. Used for AmpR expression studies.	Langley and Guest (1977)
pEc1C	Cm^R. Carries *ampRC* locus from *Enterobacter cloacae* MHN1. Template for PCR.	Nicolas et al. (1987)
pNU302	Ap^R. Carries *ampRC* locus from *Citrobacter freundii* OS60. Template for PCR.	Lindberg et al. (1985)
pBR322	Tet^R Ap^R. General cloning vehicle.	Pharmacia
pREB11	Ap^R. Carries 1.5 kb *Bam*H1 fragment of pNU302 in pBR322. Source of ligand for AmpR-DNA binding studies.	This study
pJF118EH	Ap^R. IPTG inducible *tac*-promoter expression vector. Used to express the PCR amplified *ampR* genes.	Fürste et al. (1986)
pJFEcAmpR	Ap^R. Carries PCR amplified *ampR* gene of *Enterobacter cloacae* cloned behind the IPTG inducible *tac*-promoter of pJF118EH.	This study
pJFCfAmpR	Ap^R. Carries PCR amplified *ampR* gene of *Citrobacter freundii* cloned behind the IPTG inducible *tac*-promoter of pJF118EH.	This study

Figure 3-1 *Construction of AmpR expression vectors*

(A) The Polymerase Chain Reaction (PCR) was employed to amplify the *ampR* genes of *C freundii* (*CfampR*) and *E. cloacae* (*EcampR*) without the corresponding promoter and ribosome-binding sequences (rbs). Plasmids pNU302 and pEc1c were used as templates for *CfampR* and *EcampR* respectively. Primers were designed at their 3' ends to possess complementarity with the target DNA (outlined sequences); annealing temperatures were calculated from the nucleotide content of complementary sequences. The primers contained 5' ends with ribosome-binding and/or restriction enzyme sites to facilitate cloning and expression of the amplified DNA; four additional nucleotides at the extreme 5' ends were included to improve restriction enzyme digestion. (B) The PCR amplified the *ampR* DNA without its corresponding promoter (**P**ampR) and introduced an *Eco*RI site (*CfampR*) or a *Bam*HI site (*EcampR*) followed by an efficient ribosome -binding sequence (rbs*) at the 5' ends; at the 3' ends, the *ampR* open reading frame (white box) and transcriptional terminator (**T**ampR) are followed by a *Sal*I site. The expression vector pJF118EH contains an inducible *tac*-promoter, controlled by the *lacI*^{q} gene product, and the *rrn*B transcriptional terminator which are separated by a multiple cloning sequence; selection with ampicillin is controlled by the R_{tem} ß-lactamase gene product. The PCR products and the expression vector were digested with the appropriate restriction enzymes for direct cloning of the *ampR* genes behind the *tac*-promoter (*CfampR* shown).

A.

Citrobacter freundii:

Eco RI ****rbs**********

5' ATATGAATTCAAGGAGAAATAATATGACGCGTAGCTATA 3'

Sal I

5' TATAGTCGACATCAGGCCCGCTATTGGC 3'

Enterobacter cloacae:

Bam HI ****rbs**********

5' ATATGGATCCAAGGAGAAATAATATGACGCGAAGCTATCTC 3'

Sal I

5' TATAGTCGACCCGTGGTCACCATCTGA 3'

B.

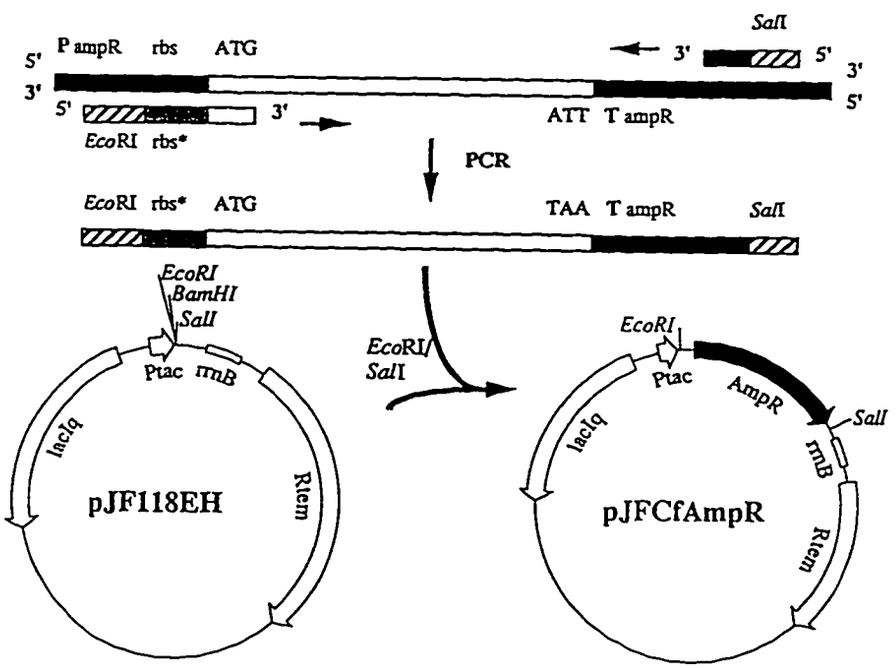


Figure 3-2 *Expression of the AmpR proteins*

E. coli strain JRG582 was transformed with plasmids pJF118EH, pJFCfAmpR, and pJFEcAmpR. Each transformant was grown to mid-logarithmic phase and the culture was made 1 mM in IPTG and allowed to grow for an additional hour. Cells were harvested, washed, and lysed in a French Pressure cell. The protein concentration of the cellular lysate was determined and 40 μg of each sample was loaded and resolved by SDS-PAGE. (A) JRG582/pJF118EH. (B) JRG582/pJFCfAmpR. (C) JRG582/ pJFEcAmpR. The positions of molecular mass standards are shown in kDa; the bands that correspond with the predicted size of the AmpR proteins are marked by an arrowhead.

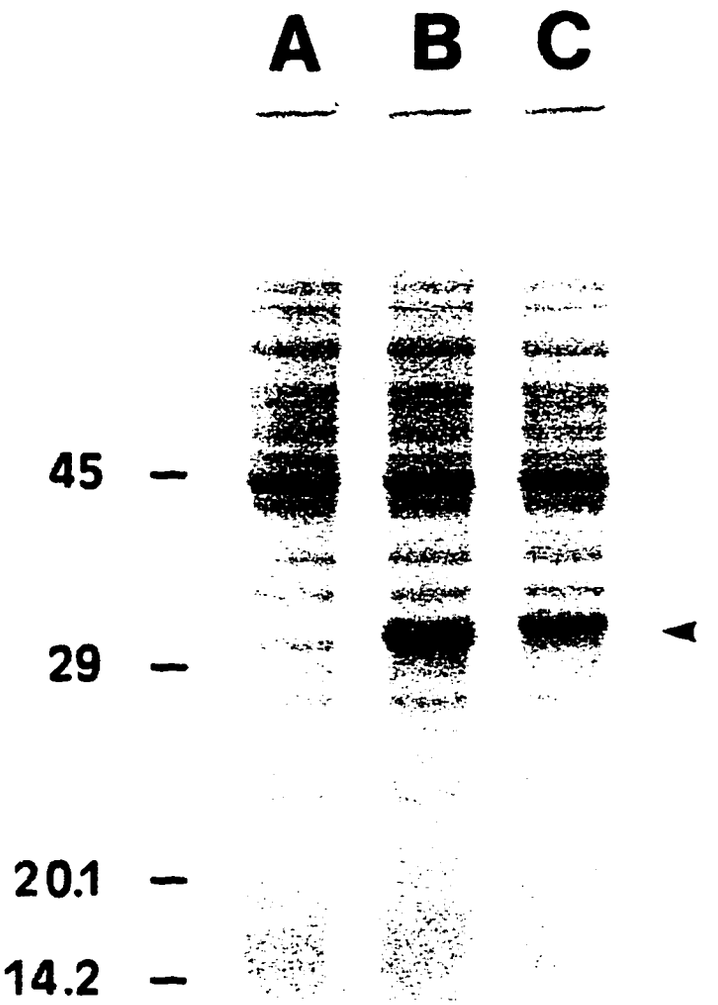

A
B
C
45
29
20.1
14.2

Figure 3-3 *Solubilization of AmpR*

AmpR extracts of JRG582/pJFCfAmpR (A-D) and JRG582/pJFEcAmpR (a-d), prepared as described in Figure 3-2, were centrifuged to precipitate insoluble material. The pellets were subjected to solubilization with 1.32 M Gdn/HCl and the extracts were centrifuged. The protein concentration of various fractions were determined and 50 μg of protein from each sample was precipitated with trichloroacetic acid to remove salts prior to SDS-PAGE analysis. (A/a) whole cell lysates. (B/b) whole cell supernatants. (C/c) pellets suspended in Gdn/HCl. (D/d) supernatants of Gdn/HCl suspension. The positions of molecular mass standards are shown in kDa; the bands that correspond with the predicted size of the AmpR proteins are marked by an arrowhead.

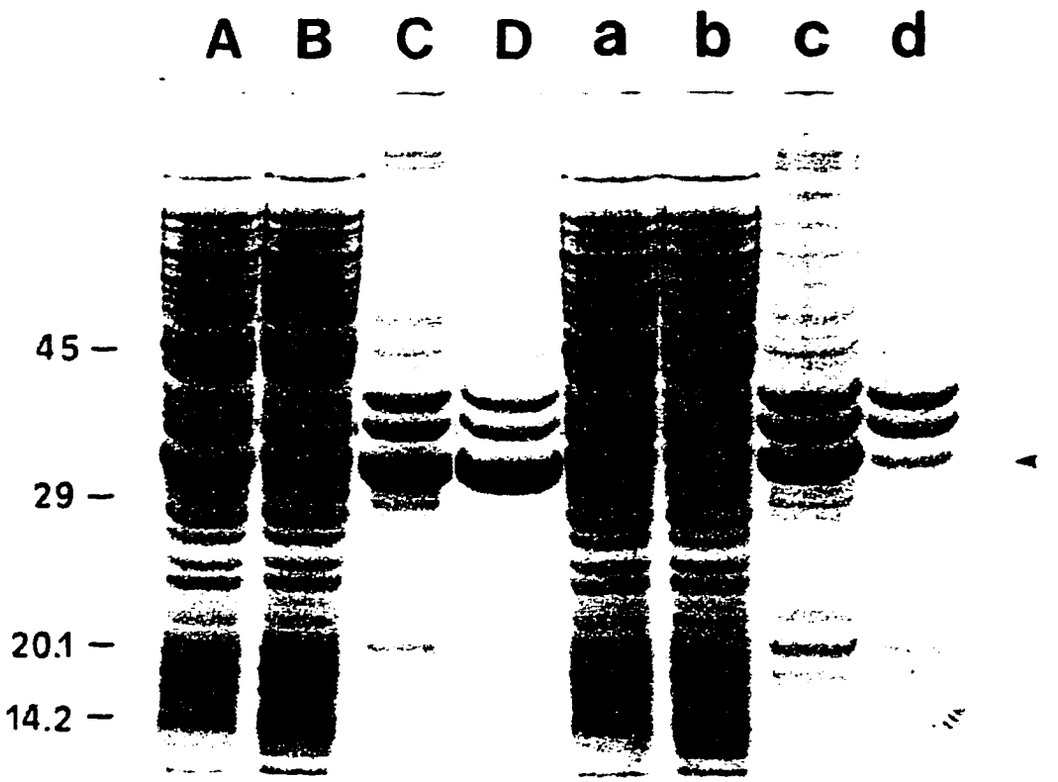

A
B
C
D
a
b
c
d
45
29
20.1
14.2

Figure 3-4 *Purification of AmpR from Citrobacter freundii*

The purification of the AmpR protein expressed in JRG582/pJFCfAmpR was monitored by SDS-PAGE analysis. (A) 40μg of protein from whole cell lysate. (B) 40μg of protein from insoluble pellet suspended in Gdn/HCl. (C) 40μg of soluble protein from the Gdn/HCl suspension. (D) 40μg of soluble protein in Buffer A-500. (E) 40μg of protein resolubilized in Buffer A-1000 after precipitation in Buffer A-50. (F, G, H) 5, 10, and 25μg of protein isolated from the peak protein fraction after Sephacryl S-200 gel-filtration in Buffer A-1000. The positions of molecular mass standards are shown in kDa; the bands that correspond with the predicted size of the AmpR proteins are marked by an arrowhead.

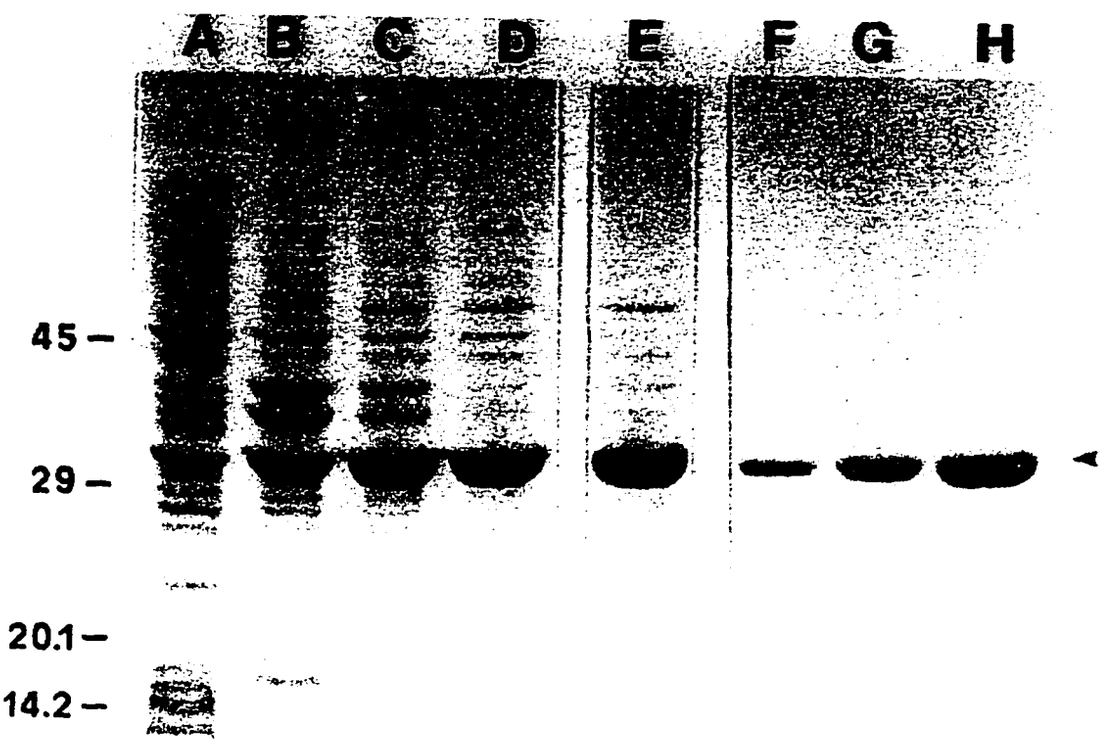

A
B
C
D
E
F
G
H
45
29
20.1
14.2

Figure 3-5 *Analytical gel-filtration of AmpR from Citrobacter freundii*

50 μg of the resolubilized CfAmpR protein in Buffer A-1000 was loaded onto a 24 mL Superose 12 HR gel filtration column operated with an FPLC system. The eluate was monitored by following A_{280}. The first broad peak eluted in the column void volume and was followed by a large monodisperse peak; both peak fractions were pooled for further analysis. The positions of molecular mass standards are shown in kDa; each of the vertical dashes are separated by a distance that corresponds to 1 mL of eluate.

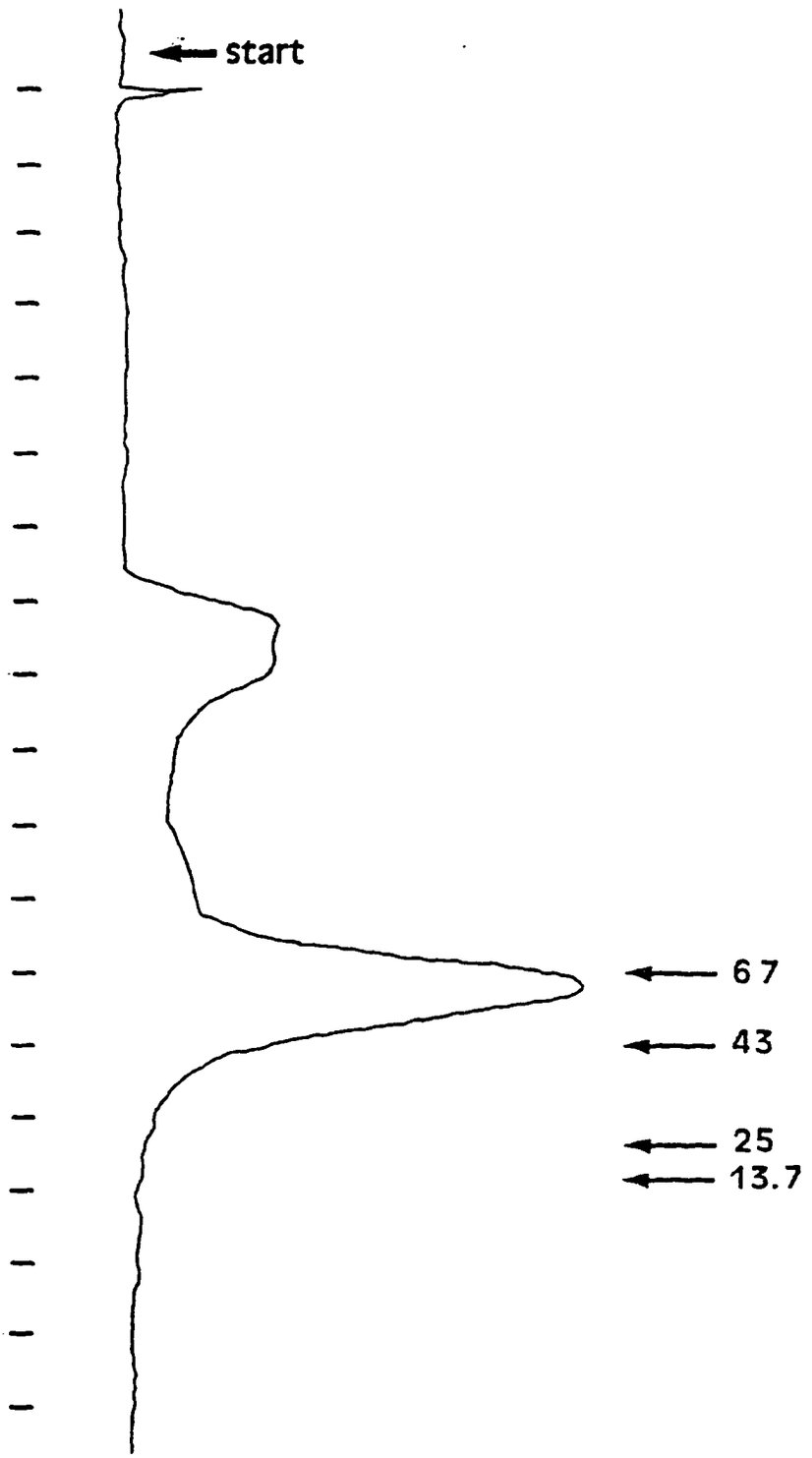

start
67
43
25
13.7

Figure 3-6 *Specific DNA-binding by purified CfAmpR*

Serial dilutions of column purified CfAmpR were made in 10 μL of Buffer C and each were mixed with 100 ng of pREB11 DNA purified from *Bam*H1 restriction digests; a 1.5 kbp *Bam*HI fragment harboring the *ampR* operator sequence and a 4.3 kbp *Bam*HI fragment derived from pBR322 respectively serve as specific and non-specific ligands to monitor DNA-binding. Reactions were equilibrated for 30 min at 25°C, separated by electrophoresis and exposed by ethidium bromide fluoresence. Quantities of AmpR used in the binding reactions are marked in ng and the positions of DNA fragments are marked in kbp; the position of the AmpR:operator complex is marked by an arrowhead.

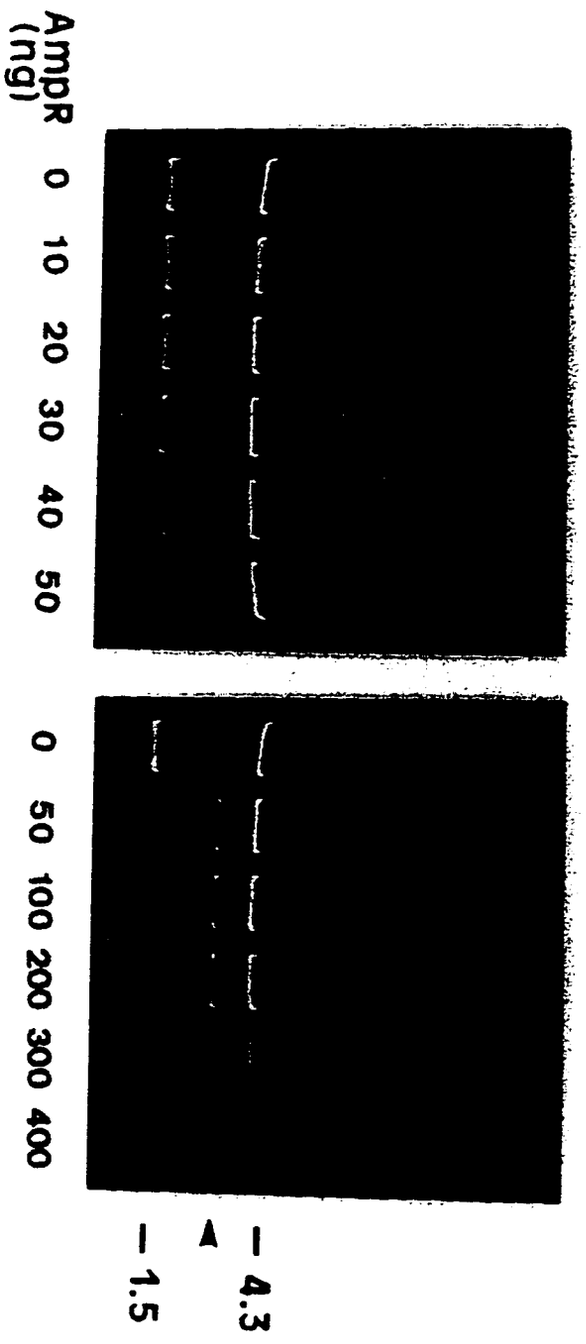
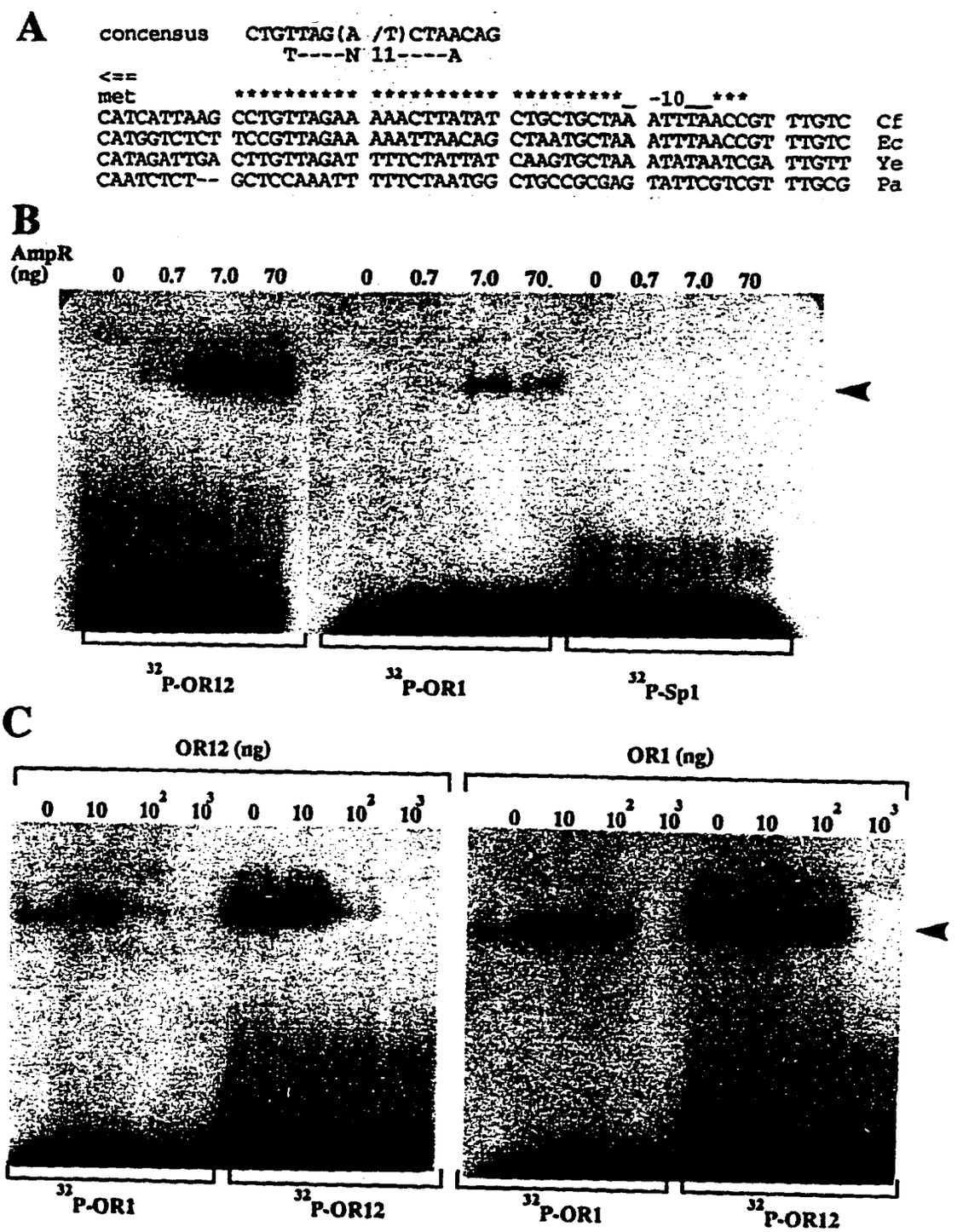

AmpR (ng)
0 10 20 30 40 50
0 50 100 200 300 400
4.3
1.5

Figure 3-7 *Binding of purified CfAmpR to a consensus half-operator*

(A) The *ampR* operator element of *Citrobacter freundii* (Cf) is overlined (***) and aligned with the corresponding sequences from *Enterobacter cloacae* (Ec), *Yersinia enterocolitica* (Ye), and *Pseudomonas aeruginosa* (Pa) (Seoane et al., 1992; Lodge et al., 1990). The -10 element of the *ampR* promoter from Ec and the *ampR* initiating Met codons are marked. A pseudo-consensus sequence corresponding to the left half of the Cf operator contains the T-N_{11}-A motif of many LysR type operators and a dyad-symmetry element flanking an A/T rich core as indicated. Elements of the pseudo-consensus half-operator are also conserved in the right half of the *ampR* operator. (B) Oligonucleotides corresponding to the top and bottom strands of the pseudo-consensus half-operator (OR1), the wild-type operator (OR12), and a non-specific sequence representing the eukaryotic *Sp1* element (Sp1) were labeled with ^{32}P by T4 kinase, incubated with purified AmpR in Buffer C in the presence of dTC/dGA, separated by electrophoresis, and exposed by autoradiography. (C) Competition assays were performed in the presence of 70ng of AmpR with excess unlabeled OR1 or OR12 as indicated. Amounts of purified AmpR and competitor DNA used in the individual binding reactions are marked in ng and the positions of the AmpR-DNA complexes are marked by an arrowhead.


A
concensus CTGTTAG(A /T)CTAACAG
T----N 11----A
<==
met ********** ********** ********* -10 ***
CATCATTAAG CCTGTTAGAA AAACTTATAT CTGCTGCTAA ATTTAACCGT TTGTC Cf
CATGGTCTCT TCCGTTAGAA AAATTAACAG CTAATGCTAA ATTTAACCGT TTGTC Ec
CATAGATTGA CTTGTTAGAT TTTCTATTAT CAAGTGCTAA ATATAATCGA TTGTT Ye
CAATCTCT-- GCTCCAAATT TTTCTAATGG CTGCCGCGAG TATTCGTCGT TTGCG Pa
B
AmpR
(ng) 0 0.7 7.0 70 0 0.7 7.0 70 0 0.7 7.0 70
32P-OR12
32P-OR1
32P-Sp1
C
OR12 (ng)
0 10 10^2 10^3 0 10 10^2 10^3
OR1 (ng)
0 10 10^2 10^3 0 10 10^2 10^3
32P-OR1
32P-OR12
32P-OR1
32P-OR12

REFERENCES

Bartowsky, E. & Normark, S. (1991) Molec. Microbiol. 5, 1715-1725.

Bishop, R.E. & Weiner, J.H. (1992) FEBS Lett. 304, 103-108

Bishop, R.E., Honoré, N., Cole, S.T., Walter, E.G. & Weiner, J.H. (1991) Program Abst. 91st Gen. Meet. Amer. Soc. Microbiol., abst. A-46.

Brennan, R.G. & Matthews, B.W. (1990) J. Biol. Chem. 264, 1903-1906.

Chang, M. & Crawford, I.P. (1989) Nucleic Acids Res. 18, 979-988.

Diekmann, S., Mazarelli, J.M., McLaughlin, L.W., von Kitzing, E. & Travers, A.A. (1992) J. Mol. Biol. 225, 729-738.

Everett, M.J., Chopra, I. & Bennett, P.M. (1990) Antimicrob. Agents Chemother. 34, 2429-2430.

Fisher, R.F. & Long, S.R. (1992) Nature 357, 655-660.

Fürste, J.P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M. & Lanka, E. (1986) Gene 48, 119-131.

Goethals, K., Montaga, M.V. & Holsters, M. (1992) Proc. Natl. Acad. Sci. USA 89, 1646-1650.

Gold, L. & Stormo, G. (1987) in: *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (Neidhardt, C. ed) pp. 1302-1308, American Society for Microbiology, Washington, D.C.

Grundström, T. & Jaurin, B. (1982) Proc. Natl. Acad. Sci. USA. 79, 1111-1115.

Henikoff, S., Haughn, G.W., Calvo, J.M. & Wallace, J.C. (1988) Proc. Natl. Acad. Sci. USA. 85, 6602-6606.

Honoré, N., Nicolas, M-.H. & Cole, S.T. (1986) EMBO J. 5, 3709-3714.

Honoré, N., Nicolas, M-.H. & Cole, S.T. (1989) Molec. Microbiol. 3, 1121-1130.

Jaurin, B., Grundström, T., Edlund, T. & Normark, S. (1981) Nature 290, 221-225.

Koo, H.-S., Drak, J., Rice, J.A. & Crothers, D.M. (1990) Biochemistry 29, 4227-4234.

Korfmann, G. & Sanders, C.C. (1989) Antimicrob. Agents Chemother. 33, 1946-1951.

Laemmli, U.K. (1970) Nature 227, 680-685.

Langley, D. & Guest, J.R. (1977) *J. Gen. Microbiol. 99,* 263-276

Leirmo, S., Harrison, C., Cayley, D.S., Burgess, R.R. & Record, T.M. Jr. (1987) Biochemistry 26, 2095-2101.

Lindberg, F., Westman, L. & Normark, S. (1985) Proc. Natl. Acad. Sci. USA 82, 4620-4624.

Lindquist, S., Lindberg, F. & Normark, S. (1989) J. Bacteriol. 171, 3746-3753.

Lodge, J.M., Minchin, S.D., Piddock, L.J.V. & Busby, J.W. (1990) Biochem. J. 272, 627-631.

Lowry, O.H., Rosenbrough, N.J., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.

Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press.

Mullis, K.B. & Faloona, F. (1987) Meth. Enzymol. 155, 355.

Nicolas, M-.H., Honoré, N., Jarlier, V., Phillipon, A. & Cole, S.T. (1987) Antimicrob. Agents Chemother. 31, 1091-1102.

Oliva, B., Bennett, P.M. & Chopra, I. (1989) Antimicrob. Agents Chemother. 33, 1116-1117.

Ottolenghi, A.C. & Ayala, J.A. (1991) Antimicrob. Agents Chemother. 35, 2359-2365.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn, Cold Spring Harbor Laboratory Press.

Schell, M.A., Brown, P.H. & Raju, S. (1990) J. Biol. Chem. 265, 3844-3850.

Seoane, A., Francia, M.V. & Lobo, J.M.G. (1992) Anitmicrob. Agents. Chemother. 36, 1049-1052.

Steitz, T.A. (1990) Quart. Rev. Biophys. 23, 205-280.

Storz, G., Tartaglia, L.A. & Ames, B.N. (1990) Science 248, 189-194.

Stoschek, C.M. (1990) Meth. Enzymol. 182, 50-68.

Tao, K., Makino, K., Yonei, S., Nakuta, A. & Shinagawa, H. (1991) J. Biochem. 109, 262-266.

Warburg, C. & Christian, W. (1942) Biochem. Z. 310, 384-421.

CHAPTER 4

Complementation of Growth Defect in an *ampC* Deletion Mutant of *Escherichia coli*[1]

Is all that we see or seem

But a dream within a dream?

Edgar Allan Poe

[1]A version of this chapter has been published. Bishop, R.E. and Weiner, J.H. (1993) *FEMS Microbiol. Lett.* **114:** 349

Introduction:

The AmpC ß-lactamase was discovered in ampicillin resistant mutants of *Escherichia coli*. The *ampC* gene is located at 94.5 minutes on the *E. coli* chromosome and encodes a class-C ß-lactamase [1]. The *ampC* promoter resides within the neighboring fumarate reductase operon (*frd*) and directs transcription through the *frd* terminator [2]; the resulting attenuator structure, designated *ampA*, coordinates *ampC* transcription with growth rate [3]. However, wild type *E. coli* cells produce insufficient quantities of the AmpC protein to confer ampicillin resistance. Rare ampicillin resistant mutants have been identified with a destabilized *ampA* attenuator [4], enhanced *ampC* promoter strength [5], or multiple tandem copies of the *ampC* gene [6]. The normal function of the growth-rate-dependent attenuator remains an enigma.

Recently, we proposed that the AmpC ß-lactamase may have an additional function as a peptidoglycan hydrolase. We suggested that AmpC cleaves the peptide bond connecting neighboring *meso*-diaminopimelate (*m*-A_2pm) residues [(L)-*m*-A_2pm-(D)-*m*-A_2pm)]; this distinct LD-endopeptidase reaction is similar to that of the LD-transpeptidase and the LD-carboxypeptidase [7]. A structural similarity between penicillin and the substrate of the DD-peptidases led to the discovery of homology between ß-lactamases and DD-peptidases [8, 9, 10]; we have demonstrated that the LD-peptidase substrate is also structurally similar to penicillin, which suggests an extended homology consistent with an LD-peptidase function for AmpC [7]. However, *E. coli* also produces a plasmid encoded ß-lactamase of class-A (RTEM) for which the sole function appears to be the detoxification of ß-lactams [11] and it is widely assumed that AmpC serves an identical function. In order to test the function of the RTEM and AmpC ß-lactamases, we investigated the growth properties of *E. coli* as a function of RTEM and AmpC expression in the presence and absence of a chromosomal *ampC* gene.

Materials and Methods:

Bacterial strains and plasmids:

Escherichia coli strains HB101 and MI1443 have been described [12]. All cloning steps were performed as described [13]. Plasmid pJFK118EH is a derivative of pJF118EH [14] in which the ampicillin resistance gene was replaced with a kanamycin resistance gene; this was achieved by blunt end ligation of the 4.6 kb *Dra*I fragment of pJF118EH with the 1.2 kb *Sma*I fragment of pUC4KIXX (Pharmacia). The 1.8 kb *Rtem* gene fragment of pBR322 [13] was isolated by *SspI/Nde*I digestion and the recessed ends filled in with Klenow DNA polymerase; blunt end ligation with the *Sma*I fragment of pBluscript KS$^-$ (Stratagene), which had been treated with calf intestinal alkaline phosphatase, generated pBluRTEM. The 1.8kb *Eco*RI/*Bam*HI fragment of pBluRTEM was ligated with the 5.8 kb *Eco*RI/*Bam*HI fragment of pJFK118EH to generate pJFKRTEM. The 1.8 kb *ApaI/Sma*I fragment of pNU302 [17] (a gift from Staffan Normark), which carries the *C. freundii ampC* gene, was ligated with the *ApaI/Sma*I fragment of pBluscript KS$^-$ to generate pBluAmpC. The 1.8 kb *KpnI/Xba*I fragment of pBluAmpC was then ligated with the 4.0 kb *KpnI/Xba*I fragment of pMS119EH [15] to generate pMSAmpC. Finally, the 1.8 kb *Eco*RI/*Sal*I fragment of pMSAmpC was ligated with the 5.8 kb *Eco*RI/*Sal*I fragment of pJFK118EH to generate pJFKAmpC.

Bacterial Growth and Gene Expression:

Transformations were performed using $CaCl_2$ as described [13]. Bacteria were grown aerobically at 37C in a gyrotory water bath shaker on LB + B1 medium (10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl , 1.1 mL 1M NaOH, 1 mL 1% thiamine, diluted to 1 L, and autoclaved). The medium was supplemented with 100 μg/mL streptomycin and 40 μg/mL kanamycin (1000X stock solutions filter sterilized and stored at -20C) prior to dispensing 10 mL of medium to individual 125 mL Klett flasks.

Isopropyl-ß-D-thiogalactopyranoside (IPTG; 100mM stock solution filter sterilized and stored at -20C) was added at a concentration of 1mM. A 5 mL stationary phase culture was grown for 24 hours and a 1% inoculum subcultured; growth was monitored with a Klett-Summerson colorimeter. Growth experiments were reproducible from day to day for the same batch of growth medium.

For protein analysis, a 25 mL culture was grown for 90 min, induced with 1 mM IPTG for 3 hours, harvested by centrifugation, washed in 10 mL of 50 mM Mops/KOH pH 7, and resuspended in 2 mL of the same Mops buffer. Cells were broken by French Pressure lysis with the 4 ml cell at 1000 psi, centrifuged at 95 000 rpm in a Beckman TL-100 ultracentrifuge using the Ti100.3 rotor at 4C for 25 min, and soluble protein measured by the method of Lowry [16]. SDS-PAGE was performed as described [17]. Low molecular mass standards from BioRad included rabbit muscle phosphorylase b (97.4 kDa), bovine serum albumin (66.2 kDa), hen egg white ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), and hen egg white lysozyme (14.4 kDa.) Purified *E. coli* RTEM and *C. freundii* AmpC were gifts from Michael James.

Results and Discussion:

Expression of ß-lactamases:

The *E. coli ampC* gene is not amenable to expression studies since the promoter is immediately followed by the transcriptional terminator of the fumarate reductase operon (*frd*) [2]. However, the *ampC* gene of *Citrobacter freundii* is separated from *frd* by a divergently transcribed regulatory gene [18]. The *C. freundii* AmpC protein is nearly 80% identical to the *E. coli* AmpC and was used to determine the molecular structure and catalytic mechanism of class-C ß-lactamases [10]. The molecular structure and catalytic mechanism of *E. coli* RTEM have also been determined [19]. Both the *C. freundii ampC*

and *E. coli Rtem* genes possess unique restriction enzyme sites between their transcriptional promoter and ribosomal-binding sequences (Figure 4-1A;) these restriction sites facilitated the replacement of the endogenous promoters with an IPTG inducible *tac*-promoter (Figure 4-1B.) Expression of RTEM and AmpC in *E. coli* HB101 and MI1443 was detected by SDS-PAGE (Figure 4-2.)

E. coli MI1443 (*ΔfrdABCD, ΔampC*) was derived from strain MM383 ($polA^{ts}$) and restored to $polA^{+}$; *E. coli* HB101 has been successfully used as a $frdABCD^{+}ampC^{+}$ strain to compare with *E. coli* MI1443 [12]. The growth properties of *E. coli* MI1443 are peculiar in that the growth rate suddenly declines at mid-exponential phase; *E. coli* HB101 exhibits typical exponential growth kinetics characterized by a straight line on the Klett scale (Figure 4-3A and 4-3B). Expression of AmpC at early-exponential phase delays the growth rate decline of *E. coli* MI1443 and inhibits the growth rate of *E. coli* HB101 (Figure 4-3C and 4-3D). We presume that adaptation of *E. coli* MI1443 to the *ampC* deletion has resulted in the observed growth defect which is relieved following AmpC production at mid-exponential phase; additionally, if growth of *E. coli* HB101 is already optimized due to the presence of a chromosomal *ampC* gene, then further AmpC production may be inhibitory. Expression of AmpC prior to exponential phase markedly inhibits the growth of *E. coli* HB101 and prevents the growth of *E. coli* MI1443 altogether (Figure 4-3E and 4-3F). It appears that AmpC production is incompatible with the transition between stationary phase and exponential phase; the growth-rate-dependent *ampA* attenuator, which down-regulates AmpC production at low growth rates, may be designed to accommodate this transition. Expression of RTEM has no effect on the growth of either *E. coli* HB101 or MI1443 under all conditions studied, suggesting that the effects of AmpC expression are specific.

Function of E. coli *ß-lactamases:*

Both RTEM and AmpC are periplasmically localized serine hydrolases which exhibit a similar ensemble of secondary structural elements [19, 10]. One difference, which may be relevant to the growth studies described here, is the observation that the RTEM and AmpC proteins differ in the temporal mode of processing during export into the periplasm; RTEM is processed post translationally, whereas AmpC is processed cotranslationally [20]. Although differential processing may explain growth inhibition by AmpC overexpression in terms of a bottleneck in the protein export machinery, it does not explain the AmpC dependent decline in growth rate at mid exponential-phase in *E. coli* MI1443.

Another difference between RTEM and AmpC concerns the constellation of hydrogen bonding interactions at the active site which determine the catalytic mechanism of ß-lactam hydrolysis; whereas RTEM utilizes a unique general base for catalysis of the acylation and deacylation steps [19], AmpC utilizes the same general base for both steps [10]. This observation may reflect the *in vitro* findings that the class-C ß-lactamase is unique in its ability to hydrolyze certain ester analogs of cell-wall peptides [21]; in some cases, the acyl-donor can be transferred to various D-amino acid acceptors in an analogous manner to the transpeptidation reactions of peptidoglycan metabolism which utilize the D-amino acid group of *m*-A_2pm as an acyl-acceptor [22]. These findings raise the possibility that AmpC may be designed to catalyze a cellular reaction related to transpeptidation in addition to the hydrolysis of ß-lactams; RTEM may be designed to function specifically as a ß-lactamase so as to avoid interference with normal cellular processes.

The nature of the cellular reaction catalyzed by AmpC, if it exists, might be discerned from the normal regulation of the *ampC* gene. Expression of the *E. coli ampC* gene is directly proportional to growth rate [3], whereas expression of the (L)-*m*-A_2pm-(D)-*m*-A_2pm peptide (LD-bridge) is inversely proportional to growth rate [23]. Although

the function of the LD-bridge is not known with certainty, it accumulates in stationary phase cells which are resistant to the lytic action of penicillin [24, 25]. The LD-bridge is thought to be formed by an LD-transpeptidase and hydrolyzed by an LD-endopeptidase in a manner analogous to the formation of the D-ala-(D)-*m*-A_2pm peptide by a DD-transpeptidase and its hydrolysis by a DD-endopeptidase (Figure 4-4)[7]. It is the LD-endopeptidase reaction that is easiest to rationalize with the growth-rate-dependent attenuator of *ampC* transcription; as optimal growth rates are achieved at exponential phase, AmpC may be required to remove LD-bridge peptides that would otherwise interfere with exponential growth. The possibility that AmpC may control LD-bridge levels is currently being studied by the analysis of cellular peptidoglycan.

Figure 4-1: *Cloning and expression of ß-lactamase genes.*

A) The transcriptional promoter (promoter) and ribosomal-binding sequences (rbs) for the *E. coli Rtem* (RTEM) [11] and *C. freundii ampC* (AMPC) [18] genes are separated by unique restriction enzyme sites as indicated (***). The positions of translational initiation codons are also indicated (met).

B) The *Rtem* and *ampC* genes were cloned into an IPTG inducible *tac*-promoter expression vector as described in Materials and Methods. Plasmid pJFKAmpC expresses the *C. freundii* AmpC ß-lactamase and pJFKRTEM expresses the *E. coli* RTEM ß-lactamase.

A:

```
                                        SspI
                                        ******      -rbs-      met
                                promoter AATATTGAA AAAGGAAGAG TATG[RTEM]
                    ApaI
                * *****                         --r bs-        met
promoter G GGCCCGGACA CCCTTTTGCT TTTAATTACG GAACTGATTT CATG[AMPC]
```

B:

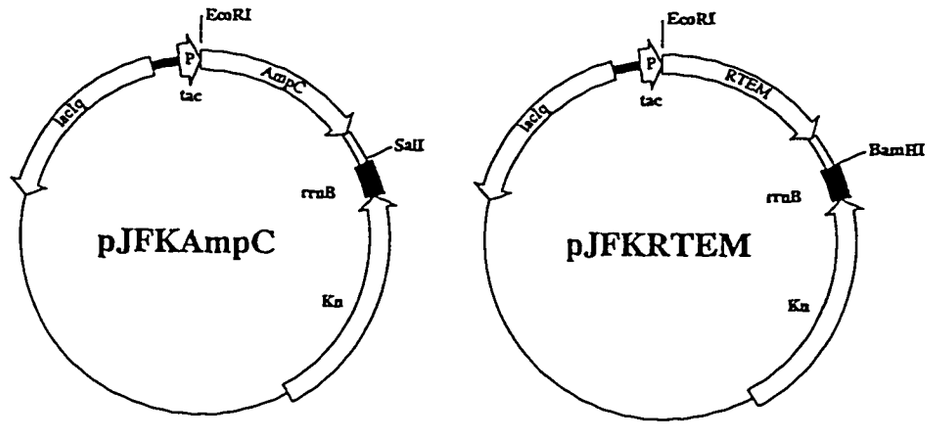
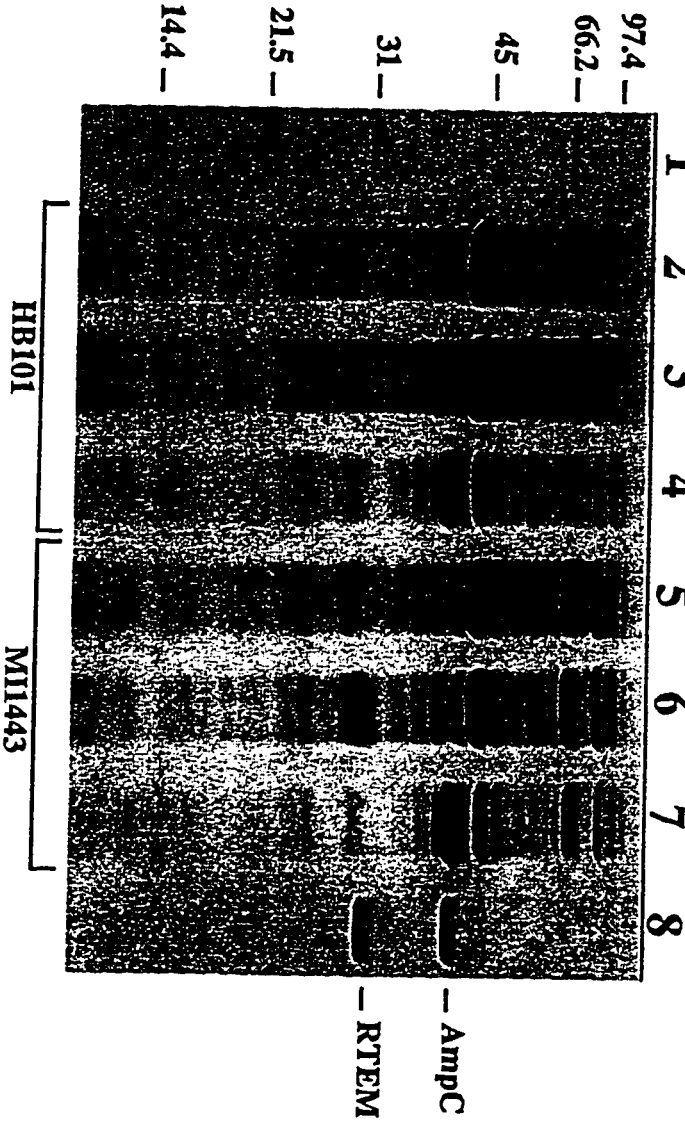


Figure 4-2: *SDS-PAGE of AmpC and RTEM proteins.*

E. coli HB101 and MI1443 were transformed with expression vector pJFK118EH (lanes 2 and 5), pJFKRTEM (lanes 3 and 6), and pJFKAmpC (lanes 4 and 7); cells were induced with IPTG and protein prepared as described in Materials and Methods. 25 μg of soluble protein from each strain was resolved on a 12% SDS-PAGE gel. Molecular mass standards are indicated in kDa (lane 1) as are the positions of purified *E. coli* RTEM and *C. freundii* AmpC (lane 8).


97.4
66.2
45
31
21.5
14.4
1 2 3 4 5 6 7 8
HB101
MI1443
AmpC
RTEM

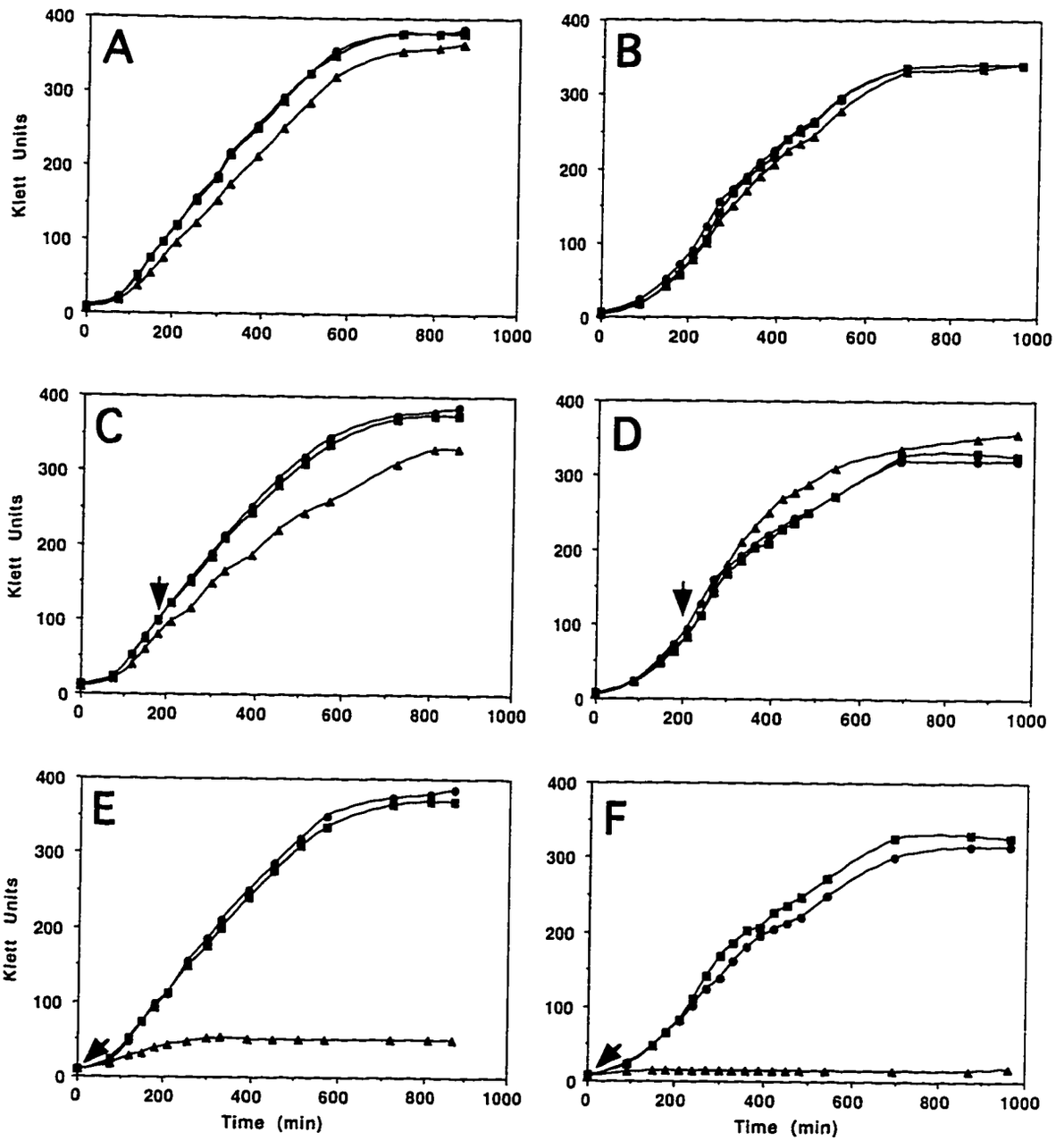


Figure 4-3: *Growth of* E. coli *HB101 and MI1443.*
Representative growth of *E. coli* HB101 (A,C,E) and MI1443 (B,D,F) transformed with plasmids pJFK118EH (■), pJFKRTEM (●), and pJFKAmpC (▲); growth was monitored with a Klett-Summerson colorimeter as described in Materials and Methods. The time of addition of IPTG is marked with an arrowhead.

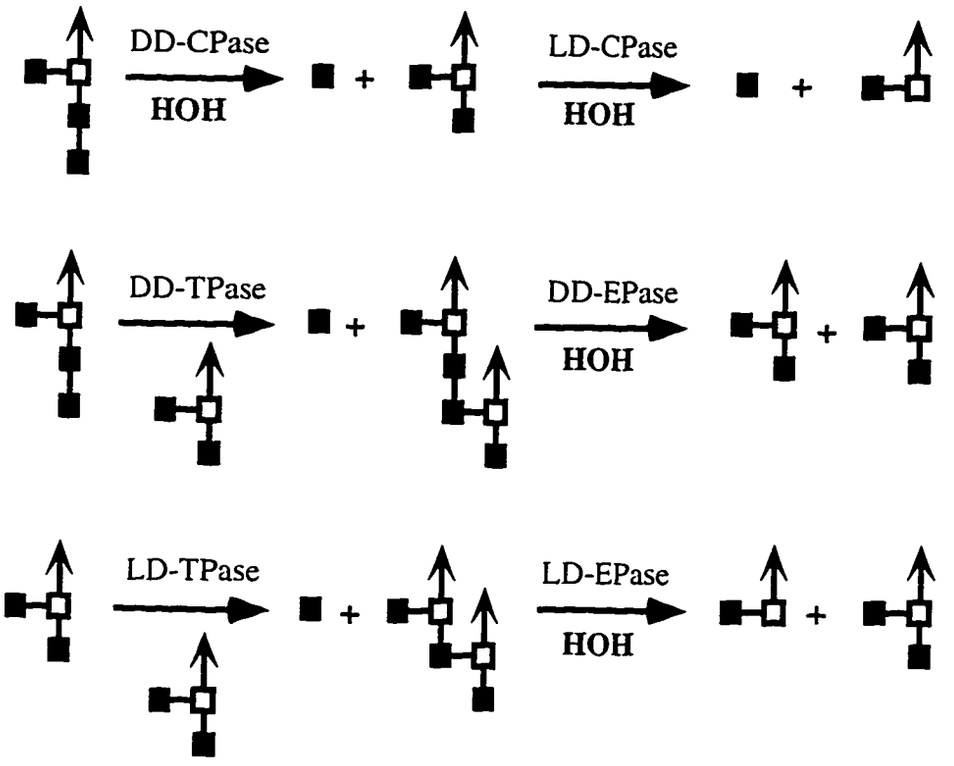


Figure 4-4 *Summary of D-alanyl carboxypeptidase metabolism.* The C-terminal D-ala residue from cell-wall peptides can be cleaved by a carboxypeptidase (CPase) or a transpeptidase (TPase) which forms a new peptide bridge; the bridge peptide can be hydrolyzed by an endopeptidase (EPase). The C-terminal and penultimate D-ala (■) residues are connected to the murein sacculus (↑) by the L centre (□) of *m*-A_2pm (■□); the ε-amino group of the D-centre of *m*-A_2pm serves as the transpeptidation acceptor using either the D-ala-D-ala peptide (DD) or the (L)-*m*-A_2pm-D-ala peptide (LD) as acyl-donor.

References:

[1] Jaurin, B., and Grundström, T. (1981) *ampC* cephalosporinase of *Escherichia coli* K-12 has a different evolutionary origin from that of ß-lactamases of the penicillinase type. Proc. Natl. Acad. Sci. U.S.A. 78, 4897-4901.

[2] Grundström, T., and Jaurin, B. (1982) Overlap between *ampC* and *frd* operons on the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. U.S.A. 79, 1111-1115.

[3] Jaurin, B., and Normark, S. (1979) *In vivo* regulation of chromosomal ß-lactamase in *Escherichia coli.* J. Bacteriol. 138, 896-902.

[4] Jaurin, B., Grundström, T., Edlund, T., and Normark, S. (1981) The *E. coli* ß-lactamase attenuator mediates growth rate-dependent regulation. Nature 290, 221-225.

[5] Olsson, O., Bergström, S., and Normark, S. (1982) Identification of a novel *ampC* promoter in a clinical isolate of *Escherichia coli.* EMBO J. 1, 1411-1416.

[6] Normark, S., Edlund, T., Grundström, T., Bergström, S., and Wolf-Watz, H. (1977) *Escherichia coli* K-12 mutants hyperproducing chromosomal beta-lactamase by gene repetitions. J. Bacteriol. 132, 912-922.

[7] Bishop, R. E., and Weiner, J. H. (1992) Coordinate regulation of murein peptidase activity and AmpC ß-lactamase synthesis in *Escherichia coli.* FEBS Lett. 304, 103-108.

[8] Samraoui, B., Sutton, B. J., Todd, R. J., Artymiuk, P. J., Waley, S. G., and Phillips, D. C. (1986) Tertiary structural similarity between a class A ß-lactamase and a penicillin-sensitive D-alanyl carboxypeptidase-transpeptidase. Nature 320, 378-380.

[9] Kelly, J. A., Dideberg, O., Charlier, P., Wery, J. P., Libert, M., Moews, P. C., Knox, J. R., Duez, C., Fraipont, Cl., Joris, B., Dusart, J., Frére, J. M., and

Ghuysen, J. M. (1986) On the origin of bacterial resistance to penicillin: comparison of a ß-lactamase and a penicillin target. Science 231, 1429-1431.

[10] Oefner, C., D'Arcy, A., Daly, J. J., Gubernator, K., Charnas, R. L.. Heinze, I., Hubschwerlen, C., and Winkler, F. K. (1990) Refined crystal structure of ß-lactamase from *Citrobacter freundii* indicates a mechanism for ß-lactam hydrolysis. Nature 343, 284-288.

[11] Sutcliffe, J. G. (1978) Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. Proc. Natl. Acad. Sci. U.S.A. 75, 3737-3741.

[12] Condon, C., and Weiner, J. H. (1988) Fumarate reductase of *Escherichia coli*: an investigation of function and assembly using *in vivo* complementation. Mol. Microbiol. 2, 43-52.

[13] Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd edn, Cold Spring Harbor Laboratory Press.

[14] Fürste, J. P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M., and Lanka, E. (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector. Gene 48, 119-131.

[15] Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992) A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the α protein of *Escherichia coli* satellite phage P4. J. Biol. Chem. 267, 13062-13072.

[16] Lowry, O. H., Rosenbrough, N. J., Farr, A. L., and Randall, R. J. (1951) Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193, 265-275.

[17] Bishop, R. E., and Weiner, J. H. (1993) Overproduction, solubilization, purification, and DNA-binding properties of AmpR from *Citrobacter freundii*. Eur. J. Biochem. 213, 405-412.

[18] Lindberg, F., and Normark, S. (1986) Sequence of the *Citrobacter freundii* OS60 chromosomal *ampC* ß-lactamase gene. Eur. J. Biochem. 156, 441-445.

[19] Strynadka, N. C. J., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K., and James, M. N. G. (1992) Molecular structure of the acyl-enzyme intermediate in ß-lactam hydrolysis at 1.7Å resolution. Nature 359, 700-705.

[20] Josefsson, L. G., and Randall, L. L. (1981) Different exported proteins in *E. coli* show differences in the temporal mode of processing *in vivo*. Cell 25, 151-157.

[21] Pratt, R. F., and Govardhan, C. P. (1984) ß-Lactamase-catalyzed hydrolysis of acyclic depsipeptides and acyl transfer to specific amino acid acceptors. Proc. Natl. Acad. Sci. U.S.A. 81, 1302-1306.

[22] Pazhanisamy, S., Govardhan, C. P., and Pratt, R. F. (1989) ß-Lactamase catalyzed aminolysis of depsipeptides: amine specificity and steady-state kinetics. Biochemistry 28, 6863-6870.

[23] Driehuis, F., and Wouters, J. T. M. (1987) Effect of growth rate and cell shape on the peptidoglycan composition of *Escherichia coli*. J. Bacteriol. 169, 97-101.

[24] Pisabarro, A. G., DePedro, M. A., and Vázquez, D. (1985) Structural modifications in the peptidoglycan of *Escherichia coli* associated with the changes in the state of growth of the culture. J. Bacteriol. 161, 238-242.

[25] Tuomanen, E., and Cozens, R. (1987) Changes in peptidoglycan composition and penicillin-binding proteins in slowly growing *Escherichia coli*. J. Bacteriol. 169, 5308-5310.

CHAPTER 5

Function of Enterobacterial AmpC ß-Lactamase

Before pronouncing a verdict of murder of a hypothesis, we should make certain that we have a corpse. A good hypothesis can survive many ugly little facts and is worth a few hundred negative experiments.

Efraim Racker

Introduction

The discovery of penicillin resistant *Escherichia coli* in the 1960's led to the identification of both chromosomal and plasmid encoded ß-lactamases (1). While the plasmid-borne RTEM ß-lactamase was clearly a virulence determinant associated with clinical ß-lactam resistance, the chromosomal AmpC ß-lactamase only conferred resistance in deregulated mutants that overproduced the enzyme. This raised suspicions that AmpC may be regulated to govern an additional physiological function (1). Since penicillin is a structural analog of peptidoglycan associated acyl-D-alanyl-D-alanine, the ß-lactamases and peptidoglycan transpeptidases were believed to be homologous. Therefore, the first *ampC* temperature-sensitive negative mutants were screened to determine if a physiological function related to peptidoglycan metabolism may explain the normal regulation of AmpC, but no positive evidence was obtained (2). AmpC is optimally produced at high growth rates when *E. coli* cells are most susceptible to penicillin induced autolysis. This led to the suggestion that *ampC* regulation may adapt the ß-lactamase to low levels of penicillins in extraintestinal environments (3). *Citrobacter freundii*, which is better adapted to extraintestinal environments than is the closely related *E. coli*, encodes a penicillin-inducible *ampC* gene. Although the *ampC* genes of both *C. freundii* and *E. coli* required mutational activation to provide clinical ß-lactam resistance, the idea that AmpC may have an additional cellular function beyond that of a ß-lactamase was largely abandoned.

However, the development of high performance liquid chromatographic procedures for peptidoglycan analysis led to the discovery of minor peptidoglycan cross bridges between neighboring diaminopimelate (A2pm) residues which were affected by growth rate (4). This finding renewed the hypothesis that AmpC may be associated with peptidoglycan metabolism and was supported by the observation of extended structural analogy between penicillin and the A2pm-A2pm bridges (5). This observation was followed by the finding that a growth defect in an *ampC* deletion mutant of *E. coli* could

be complemented by AmpC, but the result was attributable to either an effect of *ampC* gene expression or to ß-lactamase activity (6). In this study, we demonstrate that the growth defect of *ampC* mutants is independent of the ß-lactamase activity of the AmpC protein and that the A2pm-A2pm bridges of peptidoglycan are unaffected by AmpC expression. Therefore, the growth rate-dependent regulation of the *ampC* gene is probably no more than an adaptation for penicillin detoxification as suggested earlier (3).

Methods

Site directed mutagenesis was performed with the Sculptor *in vitro* mutagenesis system (Amersham) (7). Oligonucleotide synthesis and DNA sequencing was performed in the Department of Biochemistry Core Facility (University of Alberta). Cloning procedures were performed as described (8). Bacterial strains and plasmids are described in (6) and Table 5-2. Growth experiments were performed as described (6) when using *E. coli* MI1443, but derivatives of *E. coli* MC4100 were grown on Luria Bertani medium as described (8). Although it was not specified initially (9) *E. coli* DW31 carries the kanamycin resistant F plasmid RSF2001 (10) (G. Cecchini, personal communication).

Murein analysis was performed at the Max Planck Institute for Developmental Biology according to established procedures (11). Briefly, unlabeled murein sacculi were prepared by boiling cells in 4% sodium dodecyl sulfate and washing them free of detergent by repeated high speed centrifugation (60 min, 100,00 x g). Glycogen was removed by digestion with α-amylase and covalently bound lipoprotein was removed by digestion with pronase as described (11). Digestion with Cellosyl muramidase (Hoechst) yields muropeptides which must be reduced with sodium borohydride prior to separation by high performance liquid chromatography under defined conditions (11). The peaks have been isolated and identified by mass spectrometry and the column is calibrated routinely with internal standards (11). The peaks were integrated and the contribution of

the individual components specified as percentages (Table 5-1). Since some minor peaks could not be integrated accurately, the totals do not add up to exactly 100%.

1,6-anhydromuropeptides were prepared by digesting sacculi with the lytic transglycosylase purified by Astrid Ursinus according to her procedures (12). 50µL aliquots were incubated with purified *C. freundii* AmpC (provided by Michael James) at room temperature for 6hrs by addition of 10µL of 2mg/ml enzyme in phosphate buffered saline (PBS). As a control, 1,6-anhydromuropeptides were incubated with PBS under the same conditions. Samples were adjusted to 200µL with PBS, the pH adjusted to 4 with 20% phosphoric acid, then boiled for 5min, centrifuged and 200µL loaded for chromatography as performed with the reducing muropeptides (11). Since the 1,6-anhydromuropeptides have not been purified for analysis by mass spectrometry, the peaks could not be assigned. However, a qualitative assessment of the chromatograms was gained by comparison. To confirm that AmpC was active under the above conditions, a cephalosporin bioassay was performed. 30µL of 100mg/mL cephalosporin C was incubated with 20µL of 2mg/mL AmpC in PBS or with PBS alone for 2hr prior to addition of 30mL molten LB-agar in a petri dish. A loopful of an overnight culture of *E. coli* MC4100 was streaked onto the plates and an additional control plate with no antibiotic. Plates were incubated overnight at 37C.

Results

E. coli MI1443 (Δ*ampC*) carrying pJFK118EH, pJFKRTEM, or pJFKAmpC (6) were grown to mid exponential phase, induced with 1mM IPTG for 2 hours, and harvested for muropeptide analysis (11). The results are summarized in Table 5-1. No significant differences could be detected between any of the identified muropeptides. Most importantly, AmpC expression did not lead to hydrolysis of the minor A2pm-A2pm bridges as predicted earlier (5,6). This experiment demonstrates that AmpC has no effect on the murein *in vivo*, but it is possible that AmpC still affects the 1,6-

anhydromuropeptides released by the endogenous lytic transglycosylases (12). Therefore, sacculi were digested with a purified lytic transglycosylase to prepare 1,6-anhydromuropeptides (12). These were then digested with purified AmpC and resolved chromatographically. As shown in Figure 5-1, the profile of the experimental and control chromatograms are indistinguishable. In an AmpC cephalosporinase bioassay, the plate carrying AmpC digested cephalosporin C exhibited the same growth observed in the absence of antibiotic, while no growth was observed with cephalosporin C alone. These findings suggest that AmpC is active, but has no apparent activity against soluble 1,6-anhydromuropeptides *in vitro*.

In order to test the alternate hypothesis, that the growth defect in *E. coli* MI1443 is a consequence of AmpC expression and processing (6), we inactivated the enzyme by converting the nucleophilic active site serine residue to an alanine. Single stranded DNA prepared from *E. coli* TG1/pBluAmpC (6) was utilized for site directed mutagenesis with the primer named S64A (5'GCTAGGGGCGGTCAGTAAGACG3'). The recovered DNA (pBSS64A2) was confirmed by DNA sequencing to carry the oligonucleotide determined mutation that converts the active site serine 64 of AmpC to an alanine. The S64A mutant gene was moved into pJFK118EH as described (6) to create pJFKS64A2, which was transformed into *E. coli* MI1443. An IPTG induction growth experiment showed that pJFKS64A2 (Figure 5-2) gave the same growth profile observed previously with the wild type plasmid pJFKAmpC (6). The result suggests that the complementation of the growth defect in *E. coli* MI1443 by expression of AmpC is a consequence of gene expression and is independent of the biochemical activity of the enzyme.

Support for this conclusion was gained from the characterization of a smaller deletion that was created to eliminate the *frd* operon (9), but affected the 5' ends of the downstream *ampC* gene and the upstream *genX*. The deletions are outlined in Figure 5-3A and plasmids carrying individual genes encompassed by the deletions are outlined in Figure 5-3B. The pBR322 derivatives (pBRGenX and pFrd79) analyzed in *E. coli* DW33

and DW6 showed that the *frd* deletion did, in fact, result in an aerobic growth defect (Figure 5-4), but this was complemented by *genX*, a lysyl-tRNA synthetase homolog (16). The pACYC184 derivatives (pAmpAC and pAmpRC) were analyzed in *E. coli* DW31, because the *zjd*::Tn*10* allele of DW33 is incompatible with tetracycline selection. In order that *E. coli* MC4100 could be used as an isogenic control, the RSF2001 fertility factor in *E. coli* DW31 was transferred to *E. coli* MC4100 by conjugation and selection for kan^R and frd^+. The growth results (Figure 5-5A) show that the *ampC* plasmids are unable to complement the growth defect. However, a pronounced stationary phase bacteriolysis phenotype was detected. Since the *pst*I deletion of pAmpRC retained the bacteriolysis phenotype while the *pst*I deletion of pAmpAC did not (Figure 5-5B), the responsible gene or genes must lie in the uncharacterized sequence between *ampC* and *efp* (Figure 5-3B). The 2.7kb *Eco*RI/*Pst*I fragment of pAmpRC was moved into pTZ18R and pTZ19R (Pharmacia) to prepare single stranded DNA for sequencing according to the strategy outlined in Figure 5-6. The characterization of this sequence is the subject of the following chapters where important insights were gained from comparisons with the homologous sequence from *E. coli* CS520 and from the *E. coli* MG1655 genome project (17). Therefore, the sequence deduced from *Citrobacter freundii* OS60 is presented as an appendix to this dissertation.

Discussion

The AmpC ß-lactamase of *Escherichia coli* requires mutational activation to provide clinical resistance to ß-lactam antibiotics. Therefore, it has long been suspected that AmpC may possess an additional physiological function associated with the homologous D-alanyl carboxypeptidases (1,2). We demonstrate that the muropeptide profile of an *ampC* deletion strain (*E. coli* MI1443) is unaffected by expression of the AmpC protein *in vivo* and that purified AmpC does not react with 1,6-anhydromuropeptides *in vitro*. Therefore, neither intact murein nor the released 1,6-

anhydromuropeptides appear to be substrates for AmpC. While AmpC expression can complement a growth defect in *E. coli* MI1443, a Ser64Ala active site mutant of AmpC also complemented the growth defect. Therefore, the growth defect in *E. coli* MI1443 is affected by AmpC expression, but is independent of AmpC ß-lactamase activity. A smaller deletion that affects the interval between *ampC* and *genX* also exhibits an aerobic growth defect, but it is complemented by expression of GenX. Expression of AmpC failed to complement this growth defect, but resulted in a stationary phase bacteriolysis phenotype that was traced to an uncharacterized sequence in the interval between *ampC* and *efp*. Sequencing of this interval was undertaken to characterize the source of stationary phase bacteriolysis. We conclude that the only function of the AmpC ß-lactamase is to provide resistance to low levels of ß-lactam antibiotics in the natural environment, which necessitates mutational activation in order to provide resistance to the antibiotic concentrations found in clinical settings.

These findings provide strong support in favor of the hypothesis that the primary function of AmpC, and probably the only function, is to serve in detoxification of ß-lactam antibiotics (3). While the endogenous regulation does not lead to clinical ß-lactam resistance in the absence of mutational activation, the low levels of AmpC normally produced may provide resistance to the quantities of ß-lactams found in the natural environment. The fact that the RTEM ß-lactamase is regulated so as to provide quantities of enzyme that confer clinical resistance is consistent with the fact that RTEM is a clinical isolate that was probably adapted for resistance to the ß-lactam concentrations found in clinical settings. The fact that natural selection has conserved AmpC in both *E. coli* and *C. freundii* suggests that these enteric bacteria have endured exposure to ß-lactam antibiotics since the time of their common ancestor, which is more than 100 million years old (18). Whether selection for AmpC occurs in the intestinal or extraintestinal environment is not known, but the fact that *C. freundii* is better adapted to extraintestinal survival, and harbors a more complex regulation due to *ampR*, suggests

that the extraintestinal setting is the most likely. The simpler *ampA* attenuator may provide sufficient ß-lactam resistance in the commensal *E. coli*, while *Salmonella* is primarily pathogenic and has lost *ampC* altogether (19).

Despite the failure of the structural analogy between penicillin and the A2pm-A2pm peptidoglycan crossbridges (5) to predict an additional physiological function for AmpC, the hypothesis was not altogether a failure. A D-stereospecific endopeptidase with ß-lactamase activity was subsequently isolated from *Bacillus cereus* and shown to be homologous with class C ß-lactamases (20). The corollary that the inducible *ampC* gene of *C. freundii* was regulated by a peptidoglycan recycling pathway (5) also proved to have merit. It was subsequently shown that the primary pathway recycles 1,6-anhydromuropeptides (21). AmpD was shown to function as the 1,6-anhydro-MurNAc-L-alanine amidase in this pathway (22,23) and AmpG was shown to be the permease (21). The recycling of smaller murein tripeptides represents a minor component of the overall salvage process (24) and the postulated role of AmpD as a UDP-MurNAc:tripeptide ligase (5) was recently traced to a different gene product now known as Mpl (25). An elegant study, which relied on the AmpR purification procedure described in Chapter 3 (26), recently demonstrated that AmpR is repressed *in vivo* by UDP-MurNAc-pentapeptide, but this can be relieved by 1,6-anhydro-MurNAc-tripeptide (27). Therefore, the basic regulatory scheme proposed in Chapter 2 (5) compares very well with that now established (Figure 5-7), except that the system is more complex than envisaged since it depends on 1,6-anhydromuropeptides, a ß-N-acetylglucosaminidase, and dual signaling molecules. Overall, the importance of peptidoglycan recycling as a means of communicating environmental signals to intracellular regulatory mechanisms has gained wide acceptance (28). Any role of AmpE in this system remains unknown.

TABLE 5-1. Muropeptide Analysis of *E. coli* MI1443

Muropeptide[1]	pJFK118EH (%)	pJFKRTEM (%)	pJFKAmpC (%)
tri	6.86	6.40	7.74
tetra G4	1.63	1.84	1.60
tetra	34.37	33.61	35.04
di	1.64	1.80	1.24
penta	0.12	0.08	0.08
tri KR	5.67	4.76	4.77
tri Anh	0.59	0.23	0.19
tetra tri A_2pm G4	0.13	0.06	0.07
tri tri A_2pm	0.50	0.41	0.47
tetra tri A_2pm	3.52	3.33	4.05
tetra tetra G4	1.34	1.40	1.15
tetra tri	2.19	2.27	1.79
tetra penta G5	0.05	0.31	0.09
tetra tetra	26.17	28.91	27.92
tetra Anh	0.52	0.48	0.27
tri tri A_2pm KR	0.42	0.28	0.35
tetra tetra tri A_2pm	0.23	0.22	0.31
tetra tetra tri	0.17	0.18	0.19
tetra tri KR	3.70	2.66	2.93
tetra tetra tetra	2.95	3.00	2.85
tetra tetra penta	1.02	1.28	0.85
tetra tri A_2pm Anh I	0.84	0.94	0.81
tetra tri Anh I	0.58	0.56	0.41
tetra tetra tetra tetra	0.09	0.09	0.10
tetra tetra tri KR	0.47	0.48	0.42
tetra tetra Anh I	1.27	1.35	1.29
tetra tetra tri A_2pm Anh	0.03	0.06	0.03
tetra tetra tri Anh	0.11	0.12	0.10
tri tri A_2pm KR Anh I	0.07	0.24	0.12
tetra tetra tetra Anh	1.04	1.19	1.06
tri KR Anh	0.02	0.03	0.02
tetra tetra tetra tetra Anh	0.23	0.14	0.13
tetra tri KR Anh I	0.59	0.29	0.29
tetra tetra tri KR Anh I	0.11	0.08	0.08
Crosslinkage			
Ala-A_2pm	**22.46**	**23.63**	**22.23**
A_2pm-A_2pm	**2.93**	**2.79**	**3.12**

1. Abbreviations used: di, disaccharide dipeptide; tri, disaccharide tripeptide;tri KR, disaccharide tripeptide with Lys-Arg residue from lipoprotein; tetra, disaccharide tetrapeptide; penta, disaccharide pentapeptide; G4/G5, the D-alanine in position 4/5 of the peptide sidechain is replaced by glycine; Anh, 1,6-anhydromuramic acid; A_2pm, (L)-*meso*-diaminopimelyl-(D)-*meso*-diaminopimelate crossbridge.

Figure 5-1

Separation of 1,6-anhydromuropeptides by reversed phase high performance liquid chromatography. About 30µg were separated on a 3-µm Hypersil ODS column (250 x 4.6mM), using a linear gradient from methanol free 50mM sodium phosphate pH 4.31 containing 0.8mg/L sodium azide to 75mM sodium phosphate pH 4.95 containing 15% methanol (flow rate, 0.5mL/min for 135 min at 55C) as described (11). The effluent of the column was monitored at 205 nm. (A) Control sample and (B) sample incubated with purified AmpC ß-lactamase from *Citrobacter freundii*.

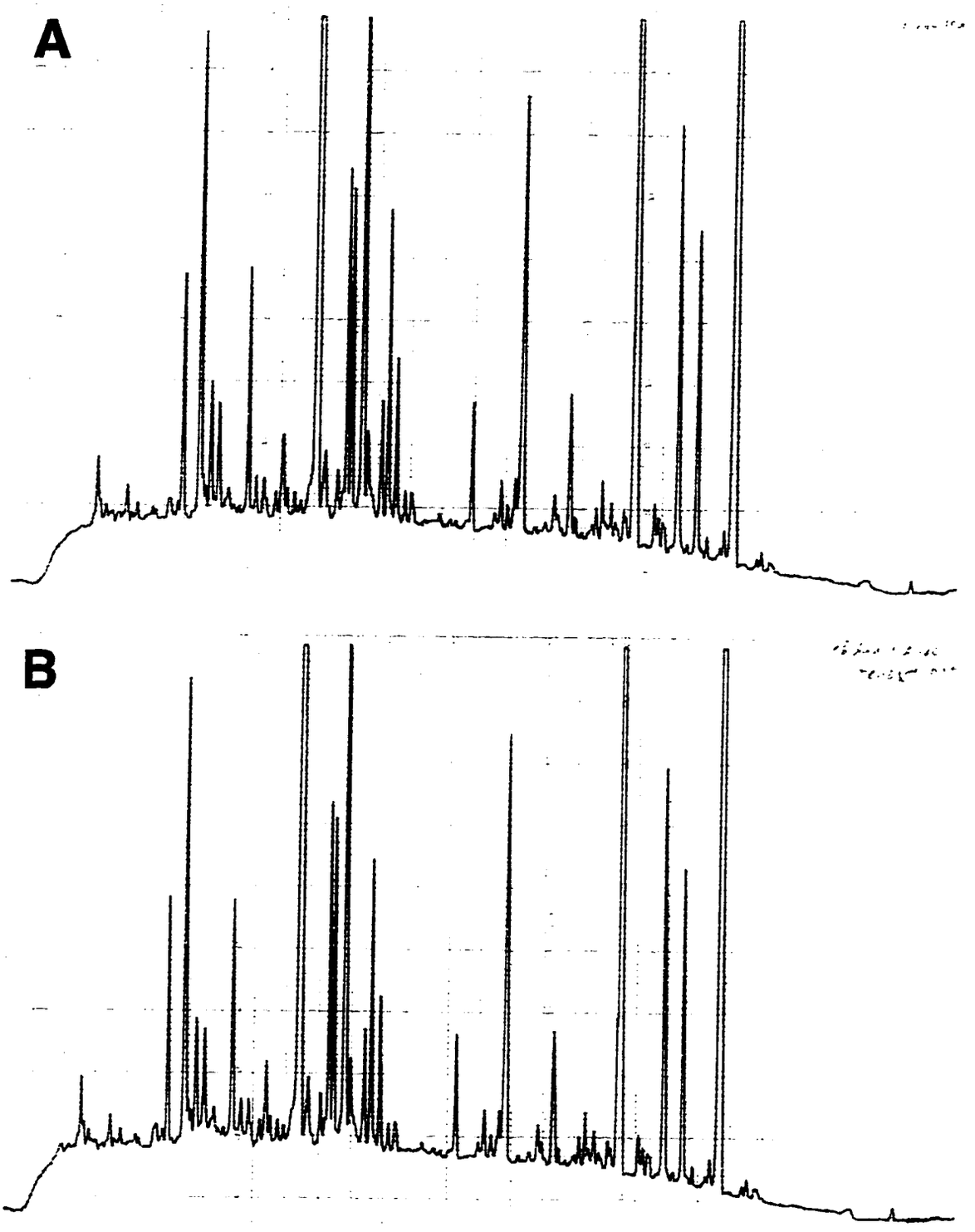

A
B

TABLE 5-2. *Escherichia coli* strains and plasmids

Strain	Relevant Genotype	Source [reference]
MC4100	F⁻, *Δ(argF-lac)U169, araD139, rpsL150, ptsF25, flbB5301, rbsR, deoC, relA1*	G. Cecchini [9]
DW6	MC4100, *zjd*::Tn*10*	G. Cecchini [9]
DW33	MC4100, *zjd*::Tn*10*, *Δ(genX-ampC)*	G. Cecchini [9]
DW31	MC4100, *Δ(genX-ampC)*, F⁺ [RSF2001(kan)]	G. Cecchini [9,10]
JG1	*supE, hsdΔ5, thi, Δ(lac-proAB)*, F'[*traD36, proAB⁺, lacIᵠ, lacZΔM15*]	Laboratory collection
MI1443	MM383, *polA⁺, rpsL, Δ(genX-efP)*	C. Condon [13]

Plasmid	Description	Source [reference]
pLC16-43	ColE1 hybrid plasmid carrying a 12.1kb fragment from the 94.5' region of *E. coli* CS520	L. Clarke and J. Carbon [14]
pNU302	7.6kb *Eco*RI fragment carrying the *ampRC* locus from *Citrobacter freundii* OS60 in pNU78	S. Normark [15]
pBluAmpC	1.8kb *Apa*I/*Sma*I fragment of pNU302 in pBSKS-	R.Bishop [6]
pBSS64A2	Ser64Ala mutant of pBluAmpC	this study
pMS64A4	1.8kb *Kpn*I/*Xba*I fragment of pBSS64A2 in pMS119EH	this study
pJFKS64A2	1.8kb *Eco*RI/*Sal*I fragment of pMS64A4 in pJFK118EH	this study
pFrd79	4.8kb *Hin*dIII fragment of pLC16-43 in pBR322	C. Condon[13]
pBSGenX	1.5kb *Pst*I/*Eco*RV fragment of pLC16-43 in pBSKS-	This study
pBRGenX	1.5kb *Bam*HI/*Sal*I fragment of pBSGenX in pBR322	This study
pAmpAC	6.1kb *Eco*RI fragment of pLC16-43 in pACYC184	This study
pAmpACΔp	*Pst*I deletion of pAmpAC	This study
pAmpRC	7.6kb *Eco*RI fragment of pNU302 in pACYC184	This study
pAmpRCΔp	*Pst*I deletion of pAmpRC	This study
pSpl18	2.7kb *Eco*RI/*Pst*I fragment of pAmpRC in pTZ18R	This study
pSpl19	2.7kb *Eco*RI/*Pst*I fragment of pAmpRC in pTZ19R	This study

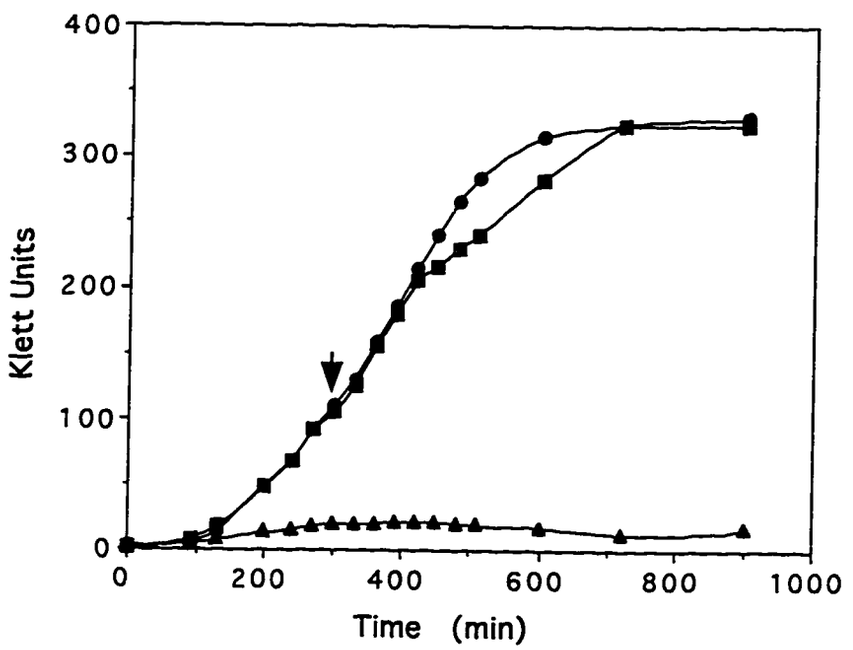


Figure 5-2 Growth of *E. coli* MI1443 transformed with plasmid pJFKS64A2. IPTG added at mid exponential phase (●) is marked with an arrowhead. IPTG added at time zero (▲) or not at all (■).

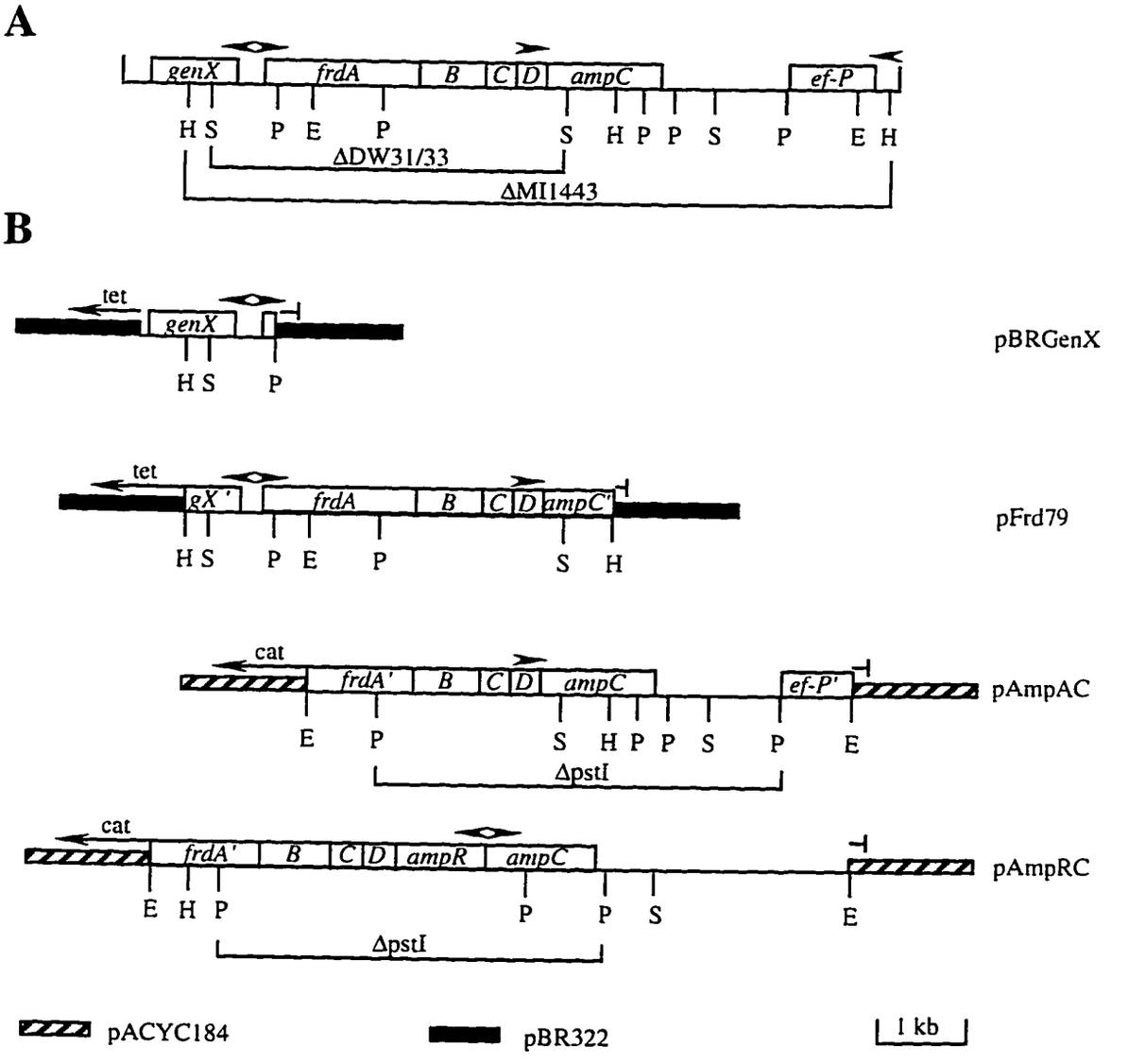


Figure 5-3 (A) Summary of the *genX-efp* interval at 94.5 minutes on the *Escherichia coli* chromosome. Deletions present in *E. coli* strains MI1443, DW31, and DW33 are indicated, and known promoter elements are marked by arrowheads. (B) Orientation of cloned genes described in Table II. Restriction sites: H, *Hin*dIII; S, *Sma*I; P, *Pst*I; E, *Eco*RI;

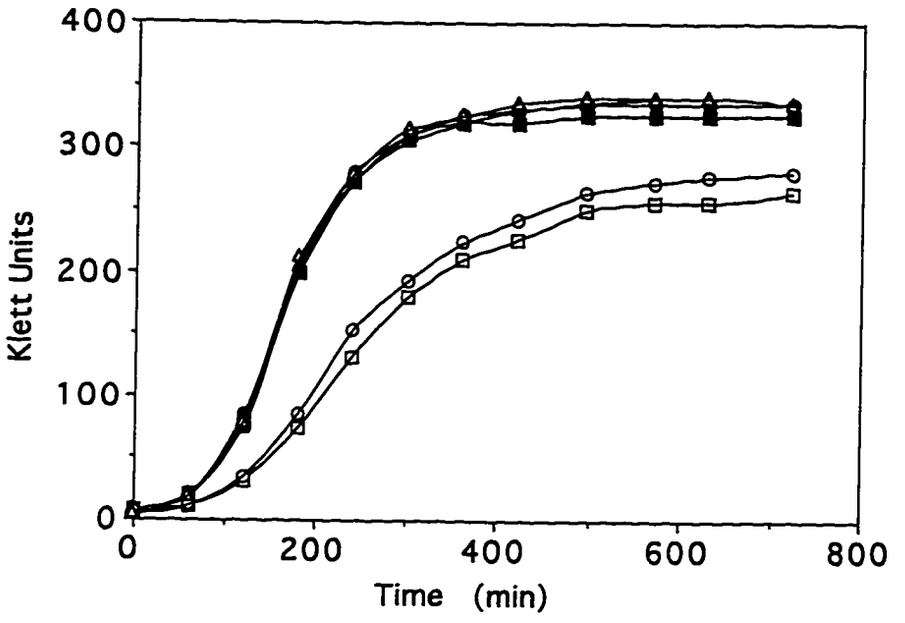


Figure 5-4 Growth of *E. coli* DW6 (solid symbols) and DW33 (open symbols) transformed with pBR322 (squares), pFrd79 (circles), or pBRGenX (triangles).

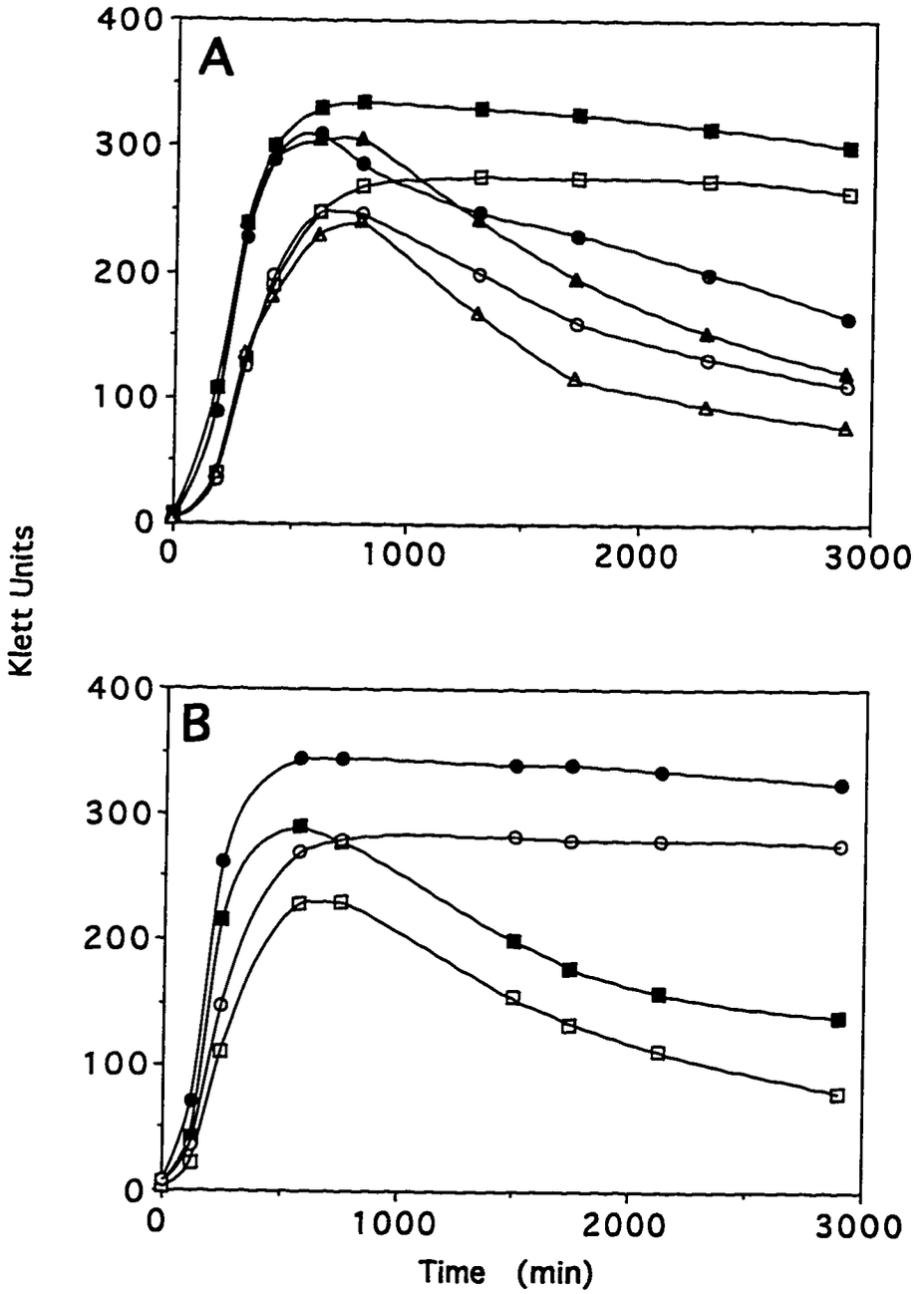
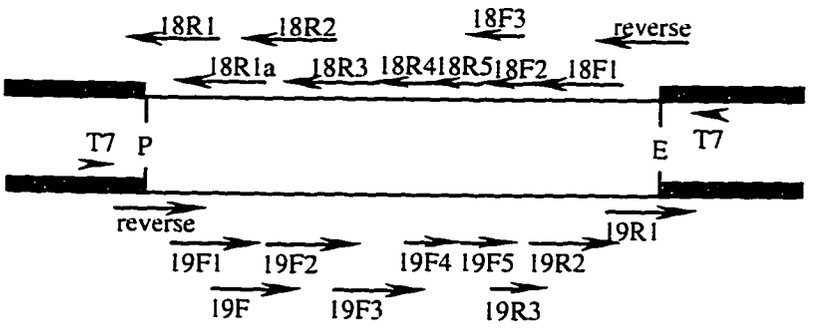


Figure 5-5 (A) Growth of *E. coli* DW31 (open symbols) and *E. coli* MC4100/ RSF2001 (solid symbols) transformed with pACYC184 (squares), pAmpAC (circles), and pAmpRC (triangles), or (B) with pAmpACΔp (circles), and pAmpRCΔp (squares).

A



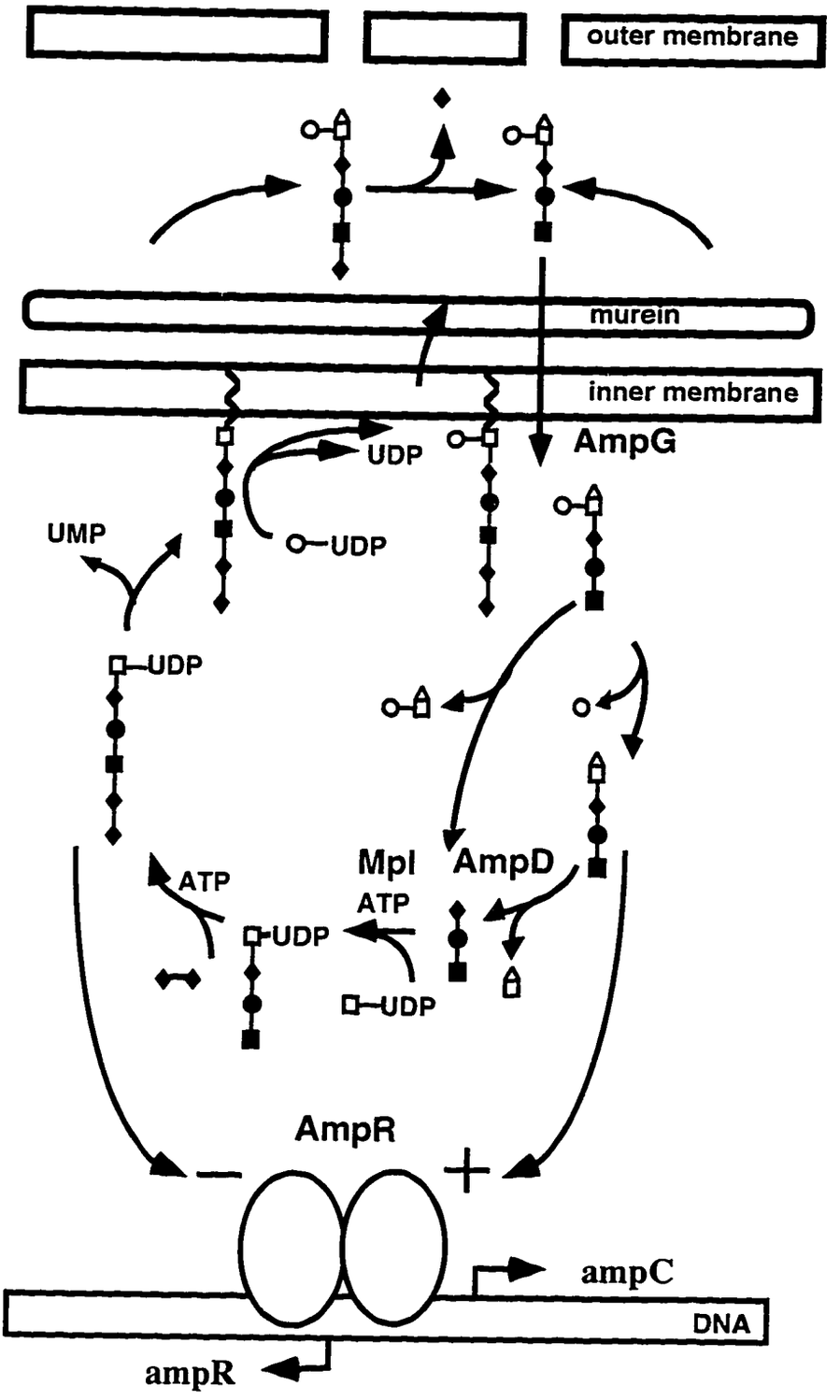
B

reverse	5' CAGGAAACAGCTATGAC 3'
18F1	5' GGCGATAACGCTAAATGG 3'
18R3	5' GCTTTTAGCTGGCAGTG 3'
18R2	5' CGCACATTACCTGTAGG 3'
18R1a	5' GCTTATTCACCTCAAACCC 3'
18R1	5' CGGTCACTTTATCCAACC 3'
19F1	5' GGTTGGATAAAGTGACCG 3'
19F	5' CCCTTCTTATGGTGGGT 3'
19F2	5' CCTACAGGTAATGTGCG 3'
19F3	5' CACTGCCAGCTAAAAGC 3'
19R2	5' GGAGTAGCATTACAAGACC 3'
19R1	5' CCATTTAGCGTTATCGCC 3'
18F2	5' GGTCTTGTAATGCTACTCC 3'
18R4	5' GGTAAAGTGGCACCGGT 3'
19F4	5' ACCGGTGCCACTTTACC 3'
18F3	5' CGCCATTTCTGGTGCAG 3'
19R3	5' CTGCACCAGAAATGGCG 3'
18R5	5' CGAACCTGAAGTGCTTCC 3'
19F5	5' GGAAGCACTTCAGGTTCG 3'

Figure 5-6 (A) Sequencing strategy of the 2.7kb *Eco*RI (E)/*Pst*I (P) fragment from *Citrobacter freundii* OS60. The T7 promoter in the pTZ18R and pTZ19R derivatives (Table 5-2) are marked by arrowheads. (B) Oligonucleotide sequencing primers as indicated above. The underlined T in 19F has no counterpart in the target DNA, but it primed sequencing successfully.

Figure 5-7

Recycling of 1,6-anhydromuropeptides in *E. coli*. Muropeptides are liberated from the murein for uptake by AmpG and then processed to the 1,6-anhydro-MurNAc-tripeptide and the free tripeptide by the combined action of a ß-N-acetylglucosaminidase and AmpD, which also acts directly on the disaccharide-tripeptide. The tripeptide produced by AmpD is coupled to the pool of nucleotide activated precursors of peptidoglycan biosynthesis by Mpl. The regulation of *ampC* is determined by AmpR, which is intrinsically an activator, but is repressed *in vivo* by UDP-MurNAc-pentapeptide. However, the 1,6-anydro-MurNAc-tripeptide, which accumulates in *ampD* mutants, can derepress AmpR. ß-lactams also lead to derepression through the combined accumulation of muropeptides and the consumption of nucleotide activated precursors. Symbols: ○ GlcNAc; □ MurNAc; ⌂ 1,6-anhydro-MurNAc; ◆ Ala; ● D-Glu; ■ *meso*-A_2pm.


outer membrane
murein
inner membrane
AmpG
UDP
UDP
UMP
UDP
UDP
ATP
Mpl
AmpD
ATP
UDP
UDP
AmpR
ampC
DNA
ampR

References

(1) Burman, L.G., Nordström, K., and Boman, H.G. (1968) *J. Bacteriol.* **96:** 438

(2) Burman, L.G., Park, J.T., Börje Lindström, E., and Boman, H.G. (1973) *J. Bacteriol.* **116:** 123

(3) Normark, S., and Lindberg, F. (1985) *Bioessays* **3:** 22

(4) Driehuis, F., and Wouters, J.T.M. (1987) *J. Bacteriol.* **169:** 97

(5) Bishop, R.E., and Weiner, J.H. (1992) *FEBS Lett.* **304:** 103

(6) Bishop, R.E., and Weiner, J.H. (1993) *FEMS Lett* **114:** 349

(7) Taylor, J.W., Ott, J., and Eckstein, F. (1985) *Nucleic Acids Res.* **13:** 8764

(8) Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press.

(9) Westenberg, D.J., Gunsalus, R.P., Ackrell, B.A., Sices, H., and Cecchini, G. (1993) *J. Biol. Chem.* **268:** 815

(10) Heffron, F., Rubens, C., and Falkow S. (1975) *Proc. Natl. Acad. Sci. USA* **72:** 3623

(11) Glauner, B. (1988) *Anal. Biochem.* **172:** 451

(12) Ursinus, A., and Höltje, J.-V. (1994) *J. Bacteriol.* **176:** 338

(13) Condon, C., and Weiner, J.H. (1988) *Molec. Microbiol.* **2:** 43

(14) Clarke, L., and Carbon, J. (1976) *Cell* **9:** 91

(15) Lindberg, F., Westman, L., and Normark, S. (1985) *Proc. Natl. Acad. Sci. USA* **82:** 4620

(16) Kong, L., Fromant, M., Blanquet, S., and Plateau, P. (1991) *Gene* **108:** 163

(17) Burland, G., Plunkett, H.J., Sofia, D.L., Daniels, F.R., and Blattner, F.R. (1995) *Nucleic Acids Res.* **23:** 2105

(18) Ochman, H., and Wilson, A.C. (1987) Evolutionary History of Enteric Bacteria, *in Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(19) Olof Olsson (1983) *Evolution of Control Sequences in Enterobacterial* ampC *β-Lactamase Genes*, Umea University Medical Dissertation No. 106, Department of Microbiology, University of Umea, Umea Sweden.

(20) Asano, Y., Ito, H., Dairi, T., and Kato, Y. (1996) *J. Biol. Chem.* **271:** 30256

(21) Jacobs, C., Huang, L.J., Bartowsky, E., Normark, S., and Park, J.T. (1994) *EMBO J.* **13:** 4684

(22) Holtje, J.V., Kopp, U., Ursinus, A., and Wiedemann, B. (1994) *FEMS Lett.* **122:** 159

(23) Jacobs, C., Joris, B., Jamin, M., Klarsov, K., Van Beeumen, J., Mengin-Lecreulx, D., van Heijenoort, J., Park, J.T., Normark, S., and Frére, J.M. (1995) *Molec. Microbiol.* **15:** 553

(24) Park, J.T. (1993) *J. Bacteriol.* **175:** 7

(25) Mengin-Lecreulx, D., van Heijenoort, J., and Park J.T. (1996) *J. Bacteriol.* **178:** 5347

(26) Bishop, R.E., and Weiner, J.H. (1993) *Eur. J. Biochem.* **213:** 405

(27) Jacobs, C, Frére, J.-M., and Normark, S. (1997) *Cell* **88:** 823

(28) Park, J.T. (1995) *Molec. Microbiol.* **17:** 421

CHAPTER 6

Stationary Phase Expression of a Novel *Escherichia coli* Outer Membrane Lipoprotein and its Relationship with Mammalian Apolipoprotein-D: Implications for the Origin of Lipocalins[1]

It doesn't matter if you fall down as long as you pick something from the floor while you get up.

Efraim Racker

[1]A version of this chapter has been published. Bishop, R.E., *et al.* (1995) *J. Biol. Chem.* **270:** 23097

INTRODUCTION

The lipocalin superfamily consists of widely distributed, primarily extracellular, eukaryotic proteins that bind and transport small hydrophobic ligands (1). The molecular structures of four lipocalins (plasma retinol binding protein (2), bilin binding protein (3), insecticyanin (4), and α-lactalbumin (5)), revealed a common structural motif that consists of an eight-stranded antiparallel ß-barrel, arranged as two stacked orthogonal sheets, with a C-terminal α-helix. Despite the common lipocalin fold, only 25-30% amino acid sequence identity exists between lipocalins of known structure (6). The cup-shaped three-dimensional structure of the lipocalins, which forms a central hydrophobic binding pocket for the ligand, is also characteristic of the fatty acid-binding proteins; these represent another recently identified protein family that also binds small hydrophobic molecules, but which by contrast, contain a ten-stranded antiparallel ß-barrel and are almost exclusively intracellular. Because of their similarities of structure and function, Flower and co-workers have proposed the classification of lipocalins and fatty acid-binding proteins into a larger structural superfamily termed calycins (1).

Although lipocalins are generally soluble proteins, apolipoprotein-D (Apo-D) was originally identified as a component of the plasma high density lipoprotein (HDL) particle, leading to the suggestion that Apo-D may transport a component of the lecithin-cholesterol acyltransferase reaction (7). Unlike the classical apolipoproteins, which are embedded in the lipoprotein surface by extended amphipathic α-helical structures, homology modeling of Apo-D against the atomic coordinates of bilin binding protein suggested that ApoD associates with the HDL particle by a hydrophobic surface loop. This modeling study also postulated that a heme-related compound may be the preferred ligand for Apo-D (8). However, Apo-D has also been identified as a progesterone- and pregnenolone-binding protein isolated from breast fluid, suggesting a role in the transport of steroid hormones in human mammary tissue (9). In the cyst fluid of women with gross cystic disease of the breast, Apo-D can exceed the concentration found in plasma by

about 1000-fold (10), and Apo-D induction by both retinoic acid (11) and interleukin-1-α (12) has been demonstrated in human breast cancer cells, suggesting that Apo-D may be a marker of hormonal alterations. Additionally, Apo-D accumulates in regenerating and remyelinating peripheral nerve, suggesting a role in lipid transport within extravascular compartments (13, 14). Like other members of the lipocalin superfamily, Apo-D appears to be able to transport a variety of ligands in a number of different contexts.

Despite the presence of lipocalins in a wide range of eukaryotic organisms, no lipocalin has ever been identified in bacteria (15). Additionally, the apolipoprotein components of plasma lipoproteins are unrelated to bacterial lipoproteins, which are anchored to membranes by a lipid-modified amino-terminal cysteine residue (16). In this report, we describe an outer membrane lipoprotein of *Escherichia coli*, which is clearly homologous to Apo-D. This protein, which we term Blc (*B*acterial *L*ipocalin), is encoded by the *blc* gene at 94.5' on the *E. coli* chromosome, immediately downstream of the *ampC* ß-lactamase operon. The *blc*-promoter is expressed at the onset of stationary growth phase under control of the *rpoS* sigma factor gene, which directs expression of genes necessary for adaptation to starvation conditions. Blc is the first lipocalin identified in a bacterial species and may provide an evolutionary link between bacterial and plasma lipoproteins.

EXPERIMENTAL PROCEDURES

Bacterial strains and growth conditions

Escherichia coli MC4100 (F^-, *Δ[argF-lac]U169, araD139, rpsL150, ptsF25, flbB5301, rbsR, deoC, relA1*) was provided by G. Cecchini (17) and RH90 (MC4100, *rpoS359*::Tn*10*) was provided by R. Hengge-Aronis (18). *E. coli* TG1 *supE, hsdΔ5, thi, Δ(lac-proAB)*, F'[*traD36, proAB*$^+$, *lacI*^{q}, *lacZΔM15*] was obtained from Amersham. Bacteria were grown aerobically in a New Brunswick shaker at 37°C in Luria Bertani

(LB) medium (19). Strains harboring plasmids or transposons were supplemented appropriately with antibiotics (ampicillin at 100 µg/ml, chloramphenicol at 42.5 µg/ml, and tetracycline at 12.5 µg/ml). Overnight cultures were subcultured at a 1% inoculum and growth monitored with a Klett-Summerson colorimeter.

Materials

DNA polymerase from *Thermus aquaticus* and restriction enzymes were obtained from BRL. Reverse transcriptase from avian myeloblastosis virus, RNaseA, bacteriophage T4 polynucleotide kinase, and glycogen were obtained from Boehringer-Mannheim, whereas RNAguard was obtained from Pharmacia. [9,10-^{3}H]palmitate (54 Ci/mmol) was obtained from Amersham Buchler, whereas [γ^{32}P]ATP (6 000 Ci/mmol) was obtained from Dupont. Globomycin was a gift from Sankyo Co., Ltd., Tokyo. 4-methylumbelliferone and 4-methylumbelliferyl-ß-D-galactoside were obtained from Sigma. All other materials were obtained from commercial sources.

DNA and protein analysis

Plasmid purifications, restriction enzyme digestions, ligations, transformations, and DNA electrophoresis were performed according to Sambrook et al. (20). Electroporation was performed with a Gibco-BRL Cell-Porator (Life Technologies, Inc., Gaithersburg, MD) according to the manufacturers instructions. DNA synthesis and sequencing was performed in the Department of Biochemistry Core Facility (University of Alberta) using an Applied Biosystems 392 DNA/RNA synthesizer and 373A DNA sequencer (Perkin Elmer, Foster City CA, U.S.A.). Protein was determined according to Lowry (21) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described by Bishop and Weiner (22).

DNA sequencing of the blc *gene from* E. coli

A 6.1 kilobase *Eco*RI fragment, corresponding to the 94.5 minute region of the chromosome from *E. coli* strain CS520, was subcloned from plasmid pLC16-43 (23) into pACYC184 (19). The resulting plasmid (pAmpAC) was propagated in *E. coli* MC4100

and used as a double stranded template for sequencing both strands of the genomic region between *ampC* (24) and *sugE* (25). The forward primer (EcPst: 5'GTCGTTGCCTGCAGTTCTCC3') hybridizes between nucleotides 148 and 167 (Fig 1), whereas the reverse primer (EcSma: 5'AACTACCAGGCTGC TGTACC3') hybridizes between nucleotides 615 and 634 (Fig 1).

Transcriptional mapping of the E. coli blc *promoter*

RNA was prepared by the modified hot phenol extraction method of Frost et al. (26) and resuspended in diethylpyrocarbonate-treated water at 10 mg/ml. The primer used in primer extension analysis (EcBlc: 5'-GGAGAACTGCAGGCAACGACC-3') hybridizes between nucleotides 147 and 167 (Fig 1). Total cellular RNA (20 μg) was annealed to 0.5 pmol of primer, labeled at the 5' end with [γ^{32}P]ATP using T4 polynucleotide kinase (19). The primer and RNA were mixed in a 30 μl volume of annealing buffer (3 M NaCl, 0.5 M Tris-HCl, pH 7.5 and 1 mM EDTA, pH 8.0), denatured at 85°C for 5 minutes, and allowed to anneal at 37°C for a minimum of 1 hour. Following annealing, the nucleic acids were precipitated with ethanol and the pellets allowed to air dry before they were resuspended in a 25 μl volume of reverse transcriptase buffer containing 0.5 mM dNTPs and 15 units RNAguard. Reverse transcriptase (20 units) was added and the reaction was incubated at 42°C for 1 hour. RNA was removed by treatment with RNaseA for 15 minutes at 37°C and the DNA was precipitated with ethanol in the presence of 0.3 M Na acetate, pH 4.8 and 10 μg glycogen. The products were separated on a 6% sequencing gel alongside dideoxy sequencing reactions using the same primer as was used for the extension reaction. Gels were dried and autoradiography was carried out at -70°C for three weeks.

Construction and analysis of a blc::lacZ *translational fusion plasmid*

A 4 kilobase *Eco*RI/*Kpn*I fragment from pAmpAC was cloned into pLacZY2 (27) to create a translational fusion plasmid (pBlcLacEK) in which the first 42 codons of *blc* are followed by *lacZ*. A 1.8 kilobase *Bam*HI/*Cla*I fragment from pMF5 (28) was cloned

into pACYC184 to create pACKatF, which carries a functional allele of *rpoS* on a vector that is compatible with pLacZY2 and pBlcLacEK; these latter two plasmids could only be transformed into *E. coli* MC4100 and RH90 by electroporation. ß-Galactosidase activity was monitored fluorometrically using 4-methylumbelliferyl-ß-D-galactoside as substrate according to the method of Youngman (29), except that activities were normalized against total protein assayed (in milligrams) rather than the optical densities of the cultures. Measurements were determined with a Sequoia-Turner model 450 fluorometer equipped with NB360 and SC430 filters.

Construction of an inducible Blc expression plasmid

Amplification of *blc* from *E. coli* by the polymerase chain reaction (PCR) was performed according to Bishop and Weiner (22) using the forward primer (EcEco: 5'ATATGAATTCAAGAGGAAACATTTATGCGCC3') for which the 18 residues at the 3' end hybridize between nucleotides 92 and 109 (Fig 6-1); the 17 residues at the 3' end of the reverse primer (EcHind: 5'TATAAAGCTTTAACT ACCAGGCTGCTG3') hybridize between nucleotides 619 and 635 (Fig 1). PCR was performed through 25 cycles of 94°C for 30 sec, 50°C for 30 sec, and 72°C for 30 sec, using pAmpAC as template. The PCR product was digested with *Eco*RI and *Hin*dIII for cloning into expression vectors pMS119EH and pMS119HE (30) digested with the same restriction enzymes; the resulting plasmids were designated pBlcEH and pBlcHE, respectively, and differ only in the orientation of the *blc* gene with respect to the inducible *tac*-promoter. The *blc* gene cloned behind the *tac*-promoter (pBlcEH) was subjected to double-stranded sequencing as indicated above using the same primers for the PCR amplification.

Membrane isolation and fractionation

E. coli MC4100 (400 ml) harboring pBlcEH or pBlcHE was grown for 2 hr and induced with 1 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) for an additional 2 hr before harvesting. Cells were resuspended in 10 ml of ice cold Buffer A (50 mM sodium phosphate pH 7.1, 10 mM $MgCl_2$), adjusted to 10 µg/ml in DNAse, and mechanically

disrupted by passage through a precooled French Press cell at 10000 psi. Unbroken cells were removed by low speed centrifugation at 4°C before the membranes were sedimented at 90 000 rpm in a Beckman TL 100.3 rotor for 30 min. Samples of the supernatant and membrane fractions, as well as the whole cell lysates, were analyzed by SDS-PAGE. The membranes were resuspended in 200 µl of ice cold Buffer A and layered on top of a discontinuous sucrose gradient composed of 0.5 ml of 60%, 1 ml of 55%, 2.4 ml of 50%, 2.5 ml of 45%, 2.4 ml of 40%, 1.4 ml of 35%, and 0.8 ml of 30% sucrose in Buffer A. Centrifugation was at 4°C for 16h at 35K in a Beckman SW41 rotor (31). The separated membranes were harvested from the centrifuge tube, and analyzed by SDS-PAGE; sucrose was removed by trichloroacetate precipitation as described by Bishop and Weiner (22).

Palmitate labeling of induced protein

For palmitate labeling, 25 ml cultures of *E. coli* MC4100 harboring pBlcEH or pBlcHE were grown for 1 hr and 10 ml transferred into growth flasks containing 50 µCi of [^{3}H]palmitate from which toluene had been evaporated over a gentle stream of N_2(g). The cultures were made 1 mM in IPTG, 100 µg/ml in ampicillin and, when necessary, 10 µg/ml in globomycin. Cells were grown an additional 1 hr, harvested, washed once with 50 mM sodium phosphate, pH 7.2 and resuspended in 120 µl of 20 mM Tris HCl pH 8 containing 1 mM EDTA and 1% SDS. Cells were incubated for 5 min at 100°C and centrifuged at 14 000 g for 10 min. The supernatant (100 µl) was added to 1 ml of ice cold acetone and kept over night at 4°C before centrifugation as above. The pellet was thoroughly resuspended in 50 µl of 1% SDS and an aliquot corresponding to about 800 000 cpm was separated by 15% SDS-PAGE and the labeled protein visualized by fluorography. Relative molecular mass was estimated using Rainbow low molecular mass standards (Amersham).

Computer methods

Sequence alignments were performed using the GCG (Genetics Computer Group, Wisconsin) FASTA and TFASTA algorithms. Protein structure was analyzed using FRODO (32) and MOLSCRIPT (33).

Nucleotide sequence accession numbers

The nucleotide sequence of the *blc* genes from *E. coli* and *C. freundii* have been deposited in the GenBank data base under accession numbers u21726 and u21727, respectively.

RESULTS

Sequence analysis of the blc *gene*

Approximately 370 base pairs of DNA that remained to be characterized between the 3' end of *ampC* and the 3' end of *sugE*, which is oriented towards *ampC* in the 94.5 minute region of the *E. coli* chromosome, was subjected to DNA sequencing. From the published sequence of *ampC* (24), we designed a forward sequencing primer (EcPst; Experimental Procedures), whereas the published sequence of *sugE* (25) allowed us to design a reverse sequencing primer (EcSma; Experimental Procedures). By utilizing the genomic DNA from *E. coli* CS520 carried on plasmid pAmpAC as a double stranded template, it was possible to unambiguously assign the sequence of both DNA strands between *ampC* and *sugE*. The 660 bp that separate *ampC* and *sugE* in *E. coli* CS520 are shown in Figure 6-1.

A single open reading frame was revealed, which specified 177 amino acid residues (19,853 Da) and exhibited a consensus prokaryotic lipoprotein cleavage site (16) predicting a mature protein of 159 residues (18,043 Da). The initiating methionine codon was separated from the rho-independent terminator of the *ampC* gene by 60 bp, from which -35 and -10 hexamers corresponding weakly to an *E. coli* σ-70 promoter could be distinguished. A reasonable ribosome binding sequence was appropriately positioned

upstream of the initiating methionine codon. The open reading frame converged upon *sugE* such that the two genes shared overlapping translational termination codons; no rho-independent terminator structure could be discerned between them. A chromosomal deletion in *E. coli* MI1443 encompasses the *blc* locus; this strain is capable of growth under both aerobic (34) and anaerobic (35) conditions, indicating that the *blc* gene is dispensable in *E. coli*.

While this manuscript was in preparation, the DNA sequence of the *E. coli* MG1655 chromosomal region from 92.8-0.1 minutes was submitted from the *E. coli* genome project under GenBank accession number u14003. The sequence from *E. coli* MG1655 confirms that determined by us for *E. coli* CS520, except for a C to G transversion at nucleotide 498; this is a silent mutation in the third position of glycine codon 132. We also determined the sequence of the *blc* homolog from *Citrobacter freundii* OS60; this DNA sequence, together with neighboring loci, will be described elsewhere. Pertinent to this study was the identification of an open reading frame specifying 177 amino acid residues and displaying 90% amino acid sequence identity with *E. coli* Blc.

Transcriptional mapping of the E. coli blc *promoter*

To verify that the *blc* gene was expressed, primer extension analysis was performed using the primer EcBlc (Experimental Procedures). A weakly expressed cDNA corresponding to a transcript whose 5' end mapped to nucleotide 80 in Figure 6-1 was detected from *E. coli* MC4100 (Figure 6-2). Autoradiography was performed for three weeks in order to detect this cDNA. The expression of the *blc* mRNA was improved if the cells were allowed to grow into stationary phase.

E. coli MC4100 was recently reported to be distinct from many other common strains of *E. coli* by virtue of having a functional allele of *rpoS*, which controls a program of gene expression induced under starvation conditions and at the onset of stationary phase (36, 37). In order to determine if the accumulation of the *blc* transcript in

stationary phase was determined by *rpoS*, we also performed primer extension analysis using *E. coli* RH90 (MC4100 *rpoS359*::Tn*10*). The *blc* transcript was not detected in *E. coli* RH90 under all observed growth phases (Figure 6-2), indicating that the *blc* gene belongs to the stationary phase regulon controlled by *rpoS*, and suggesting that the Blc protein may serve a function that contributes to the adaptation of cells to starvation conditions. Although there is no clearly defined consensus sequence for *rpoS*-dependent promoters, they are generally similar in structure to those controlled by σ-70 (38, 39).

Monitoring activity of a blc::lacZ *translational fusion* in vivo

To determine if the primer extension results were valid *in vivo*, a *blc::lacZ* translational fusion was created from pLacZY2, which was developed to exhibit low background ß-galactosidase activity, thus permitting the analysis of weakly expressed genes (27). The results (Table 6-I) demonstrate that activities expressed from the fusion plasmid (pBlcLacEK) were only significantly above those expressed from pLacZY2 when a functional *rpoS* allele was present, indicating that σ-70 has no detectable effect on the *blc* promoter. A slight 1.5 fold increase above background activity was determined in exponential phase cultures (200 minutes of growth) in the presence of the chromosomal *rpoS* gene from *E. coli* MC4100, and this was matched by a 1.4 fold increase in the presence of the plasmid encoded *rpoS* gene (pACKatF) harbored by *E. coli* RH90. However, in stationary phase cultures (overnight growth), the chromosomal *rpoS* gene provided a 2.5 fold increase above background activity, whereas the plasmid encoded *rpoS* gene provided a 5.3 fold increase. These results confirm that *blc* is normally expressed only very weakly in *E. coli* and that this expression is optimized in stationary phase under the control of *rpoS*.

Controlled expression of Blc protein

To express the Blc protein, the *blc* gene from *E. coli* was cloned behind the IPTG-inducible *tac*-promoter in the expression vector pMS119EH to create pBlcEH. The *blc* gene was also cloned in the reverse orientation in pMS119HE to create pBlcHE, which

serves as a negative control. The gene was amplified by PCR using primers designed to incorporate an *Eco*RI restriction enzyme site and improved ribosome binding sequence at the 5' end (EcEco; Experimental Procedures), and a *Hin*dIII restriction enzyme site at the 3' end (EcHind; Experimental Procedures). The DNA sequence of the PCR product in pBlcEH was identical to that of the template, except for a G to A transition at nucleotide 516; this is a silent mutation in the third position of leucine codon 138. Blc was induced in *E. coli* MC4100 and protein divided into cellular lysate, soluble, and membrane fractions. Samples analyzed by SDS-PAGE are shown in Figure 6-3. A doublet of bands migrating with molecular masses of 16 and 18 kDa in the Blc expressing lanes were visible in the cellular lysate and the membrane fractions. This finding is consistent with an accumulation of a membrane protein which is being processed to a mature species by the removal of an 18 amino acid signal peptide and the addition of 700 Da, the approximate molecular mass attributed to lipid modification of prokaryotic lipoproteins. The two unique bands in the Blc expressing lanes migrate slightly faster than their expected positions of 18 and 20 kDa for the mature and precursor species of Blc, respectively. This anomalous migration may result from lipid modification of Blc, which would promote binding of SDS and facilitate migration in SDS-PAGE.

Subcellular localization of Blc protein

To localize the two Blc species in the Gram negative cell envelope of *E. coli*, membranes were fractionated by sucrose density gradient centrifugation. Light, heavy, and medium fractions were harvested; these represent the cytoplasmic membrane, outer membrane, and adhesion zones between outer and cytoplasmic membranes, respectively (31). The band corresponding to the Blc precursor was localized in the cytoplasmic membrane fraction, whereas the band corresponding to mature Blc was localized in the outer membrane fraction, and both species were equally distributed among the fraction of adhesion zones (Figure 6-4). This result suggests that the Blc precursor accumulates in the cytoplasmic membrane, and that mature Blc is targeted to the outer membrane.

Palmitate labeling of the Blc protein

Further evidence of the lipoprotein nature of the Blc protein was obtained by expressing the protein in the presence of [^{3}H]palmitate. A major band near 16 kDa was detected by fluorography after SDS-PAGE only when Blc was expressed (Figure 6-5). Addition of globomycin (40), an inhibitor of the lipoprotein signal peptidase (signal peptidase II), inhibited the formation of the 16 kDa band concomitantly with the formation of an 18 kDa band (Figure 6-5); this latter species likely represents the diacylglyceryl-prolipoprotein form of Blc. The heavy band near 5 kDa corresponds to the murein lipoprotein for which a larger diacylglyceryl-prolipoprotein form also appears in the presence of globomycin (16).

Homologies between Blc and Apo-D

A sequence alignment using TFASTA (Genetics Computer Group, Wisconsin) revealed that the *E. coli* and *C. freundii* Blc species exhibited homology with Vlp, an uncharacterized open reading frame derived from *Vibrio cholerae* (Table 6-2). Additionally, the bacterial proteins scored between 26 and 33% amino acid sequence identity over their entire length with three species of mammalian Apo-D. Lower level similarities were detected with a number of other lipocalins. The initial alignments were optimized by comparison of the primary structure of Blc with the tertiary structure of human Apo-D using FRODO (32). The resulting alignment is shown in Figure 6-6.

DISCUSSION

We have identified a novel bacterial lipoprotein (Blc), which exhibits homology with a eukaryotic lipocalin (Apo-D). Blc is optimally expressed in stationary phase and is under the control of the RpoS σ-factor global regulator. In the natural environment, bacteria spend the majority of their existence in the stationary phase and the expression of genes during stationary phase is of considerable interest. Bacteria have developed sophisticated mechanisms to survive starvation for prolonged periods of time and it has

been proposed that proteins synthesized at the late stages of growth are important for the survival of the organism (41, 42). It is now apparent that several *E. coli* lipoproteins are expressed under stationary phase conditions. A starvation inducible lipoprotein (Slp) was recently shown to be expressed in stationary phase cultures independently of *rpoS* (43), whereas stationary phase expression of the OsmB lipoprotein was shown to require *rpoS* (44). The *rpoS*-dependent expression of the *blc* gene suggests that Blc, like the other stationary phase lipoproteins, may serve an important starvation response function in *E. coli*.

A Blc-like protein has been found in *E. coli, C. freundii* and *V. cholerae* and it presumably exists in other *Enterobacteriaceae*. Although an *E. coli* strain deleted for the *blc* gene grows normally in the laboratory, *blc* may offer some advantage in the pathogenic or natural environment. Based on the hydrophobic ligand binding capacity of ApoD, it seems likely that Blc also functions to bind a hydrophobic ligand. The Blc mRNA is very poorly expressed, suggesting that the normal level of Blc in the outer membrane is low and that Blc is not simply sequestering hydrophobic ligands. Most outer membrane lipoproteins are oriented towards the periplasm and we presume that Blc serves to capture its hydrophobic ligand within the periplasm, although we have not eliminated the possibility that the protein serves as an external receptor molecule. Studies are in progress to characterize the orientation of Blc.

The lipocalin proteins are composed of a common structural motif that consists of an eight-stranded antiparallel ß-barrel, arranged as two stacked orthogonal sheets, with a C-terminal α-helix. A large proportion of lipocalin residues are exposed to the hydrophobic interior and reside in the central ligand binding cavity. It has been suggested that different lipocalins change these residues to accommodate different ligand specificities, thus explaining the rather low levels of primary structural homology despite the common lipocalin fold (6). In this context, the 26-33% amino acid sequence identity that exists between the three species of Apo-D and the three bacterial lipocalins must be

regarded as highly significant (Table 6-2 and Figure 6-6). Molecular modeling of Apo-D against the atomic coordinates of Bilin Binding Protein (8), which together share only 27% amino acid sequence identity, predicted a disulfide bonding pattern in Apo-D that has since been confirmed biochemically (45). Additionally, the Apo-D model led to the identification of a hydrophobic surface loop between ß-strands 7 and 8, which was postulated to mediate the physical interaction between Apo-D and HDL (Figure 6-7). It has since been shown that a lone cysteine residue located immediately adjacent to the hydrophobic loop in human Apo-D forms an intermolecular disulfide bond with the HDL associated apolipoprotein-AII, providing strong support for the proposed interaction (45). Interestingly, the aforementioned cysteine residue is absent in the rodent Apo-D species (Figure 6), which is consistent with the observation that rat Apo-D is not found to a significant extent on plasma lipoproteins (13).

A 7 amino acid deletion in the bacterial lipocalins appears to have eliminated the corresponding hydrophobic surface loop found in the Apo-D species (Figure 6-6). Additionally, a glycine and a proline residue are suitably positioned at the leading edge of the deleted loop to accommodate a ß-hairpin turn. A lone cysteine preceding this putative turn structure is evident in the *E. coli* and *C. freundii* Blc proteins, but not in the *V. cholerae* Blc, potentially implicating a species specific disulfide between Blc and itself or another component of the bacterial cell envelope. Two intramolecular disulfide bonds in the Apo-D species are formed by four additional cysteine residues, which are absent in the three bacterial lipocalins. The only other cysteine residue in the Blc proteins is located at the lipoprotein processing site near the amino terminus. The *in vivo* localization of *E. coli* Blc in the outer membrane is consistent with the presence of a serine residue at the +2 position of the mature protein (46).

As Blc represents the first lipocalin to be characterized from a bacterium, it raises important questions regarding the origin of this burgeoning class of proteins. Blc is unique among bacterial lipoproteins in that it is a lipocalin, whereas Apo-D is unique

among lipocalins in that it is a plasma lipoprotein. Did the primordial lipocalin originate in bacteria, where it served a function in the Gram-negative cell envelope before it was acquired by eukaryotes and adapted to a number of functions in multicellular organisms, or did the lipocalins originate firstly in eukaryotes from where they were adapted, at least in the case of Apo-D, to function in the Gram-negative cell envelope? What should be clear from this study, is that the function of Apo-D either originates from or has been adapted toward a basic function in the cell envelope of *E. coli*. In this regard, it is remarkable that Blc and Apo-D are both found anchored in asymmetric bilayers; the inner leaflet of the Gram-negative outer-membrane and the plasma exposed surface of mammalian lipoproteins are two rare examples of non-phospholipid bilayer membranes in biology. Perhaps the still undefined physiological functions of Blc and Apo-D may prove to be more closely related than otherwise anticipated. The *rpoS*-dependent activation of the *blc* gene at the onset of stationary phase suggests that the Blc protein may somehow serve in adaptation to starvation conditions. We hope that further analysis of Blc structure and function will be relevant to Apo-D and its pathology.

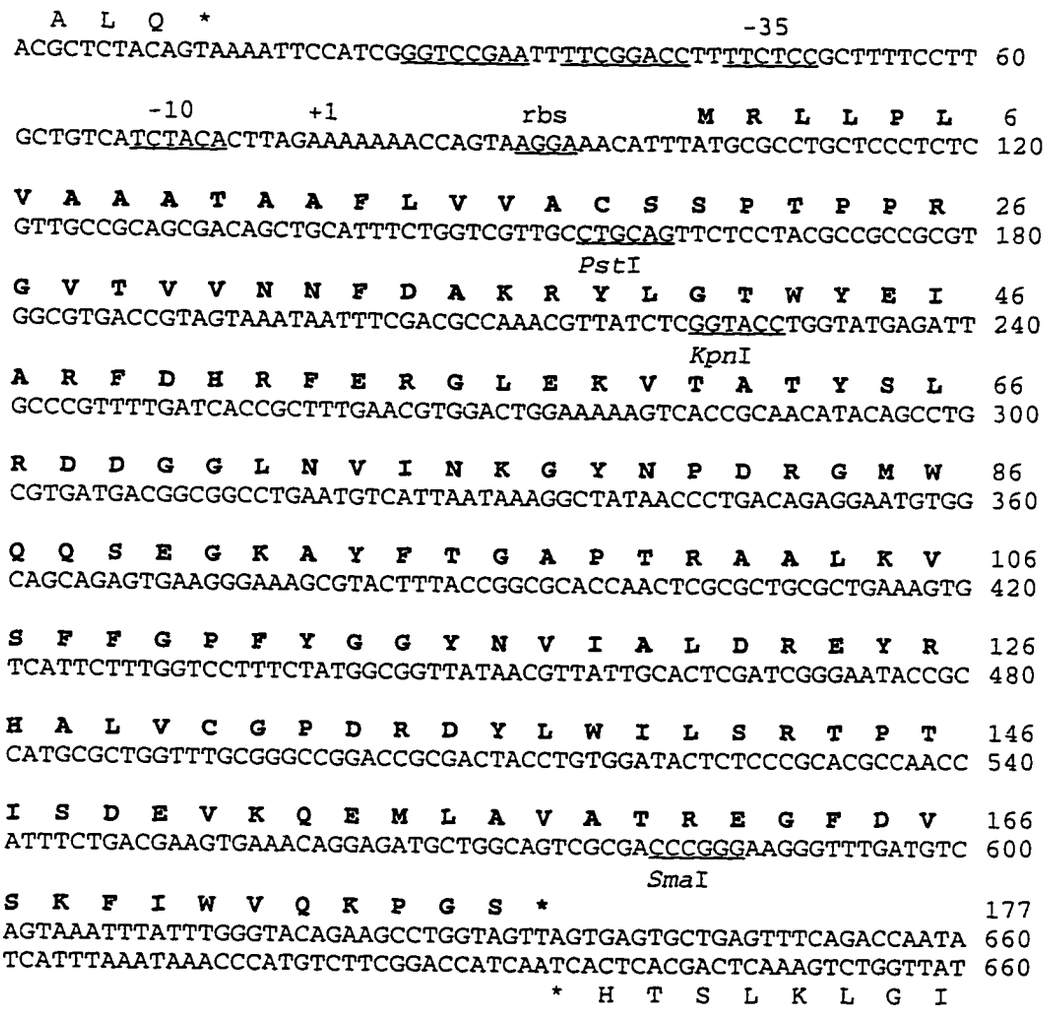


Figure 6-1 *Nucleotide sequence of the* blc *gene from* E. coli *CS520*

The sequence of the *blc* gene from pAmpAC extends over 660 base pairs beginning with the final four codons of the *ampC* gene and its rho-independent terminator (underlined), followed by the -35 and -10 regions of the *blc* promoter, the *blc* transcriptional start (+1) and ribosomal binding (rbs) sites, through the 177 amino acid residues of the Blc open reading frame (bold type), and ending with the final 9 codons of the convergent *sugE* gene. The *Pst*I, *Kpn*I, and *Sma*I restriction enzyme sites are underlined and the cysteine residue at the lipoprotein cleavage site is outlined.

Figure 6-2 *Transcriptional mapping and* rpoS *control of the* blc *promoter*

The deduced sequence of the *blc* promoter region derived from the sequencing lanes labeled G, A, T, and C is shown vertically on the left. The primer extension reactions were performed with total cellular RNA isolated from *E. coli* strains MC4100 (*rpoS*$^+$) and RH90 (MC4100, *rpoS359*::Tn*10*). RNA samples were isolated at mid-exponential (1), late-exponential (2), early-stationary (3), and stationary (4) growth phases, which were reached after 200, 270, 340, and 540 minutes of growth, respectively. The position of the cDNA corresponding to the *blc* transcription start site (+1) is marked by an arrowhead.

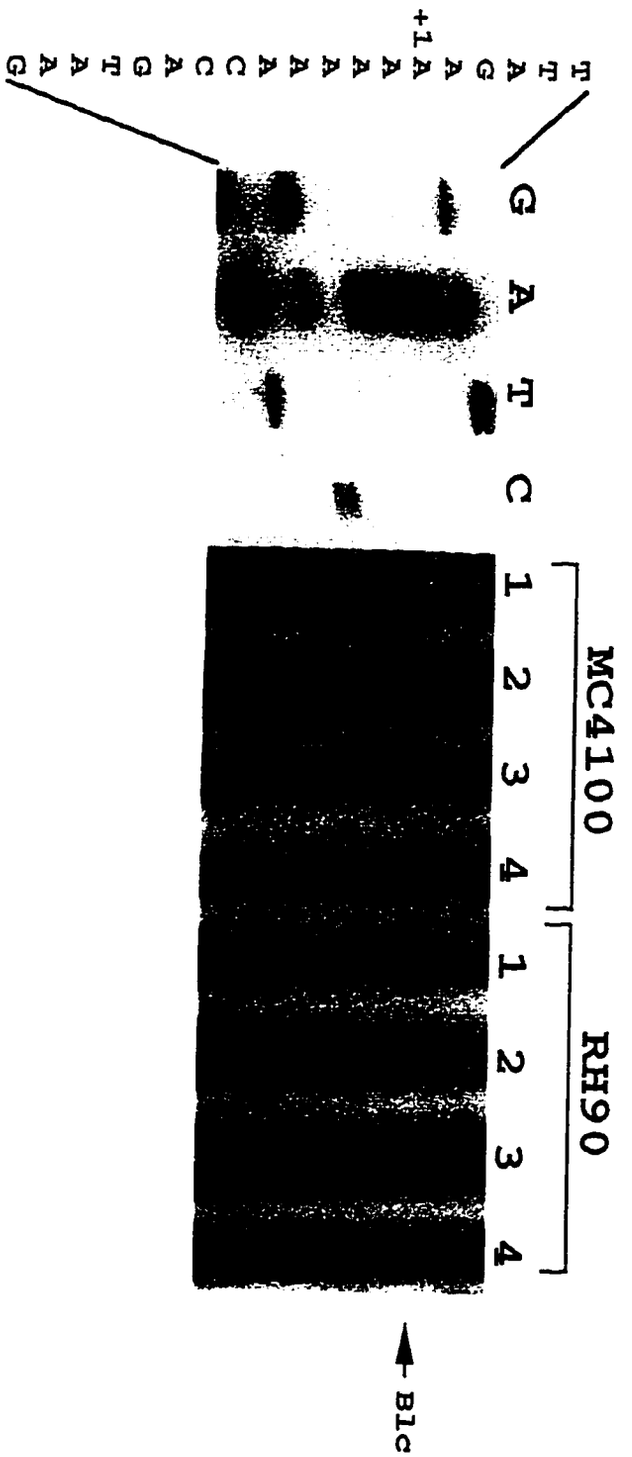

T
T
A
G
A
A
A
A
A
A
A
+1A
A
A
A
A
C
C
C
A
G
T
G
A
A
A
A
G
G
A
T
C
1
2
3
4
1
2
3
4
MC4100
RH90
Blc

Table 6-1

ß-Galactosidase activities expressed from a blc::lacZ translational fusion plasmid

ß-Galactosidase specific activities were monitored in exponential phase and stationary phase cultures from a chimeric *blc::lacZ* translational fusion plasmid (pBlcLacEK) and the promoterless control plasmid (pLacZY2) in the presence and absence of a functional *rpoS* allele, which was carried either on pACKatF or on the chromosome of *E. coli* MC4100. Results are the mean values (± standard deviation) of three independent fluorometric assays using 4-methylumbelliferyl-ß-D-galactoside as substrate.

Strain	Plasmid	ß-Galactosidase Specific Activity (pmol/min·mg)	
		Growth Phase	
		Exponential	Stationary
MC4100 (*rpoS*$^+$)	pLacZY2	13 ± 1	13 ± 2
MC4100 (*rpoS*$^+$)	pBlcLacEK	19 ± 1	33 ± 2
RH90 (*rpoS*$^-$)	pLacZY2	15 ± 1	14 ± 2
RH90 (*rpoS*$^-$)	pBlcLacEK	16 ± 1	16 ± 1
RH90 (*rpoS*$^-$)	pACKatF (*rpoS*$^+$) pLacZY2	16 ± 3	15 ± 1
RH90 (*rpoS*$^-$)	pACKatF (*rpoS*$^+$) pBlcLacEK	23 ± 2	80 ± 10

Figure 6-3 *Overexpression and membrane localization of the Blc protein*

12% SDS-PAGE analysis of *E. coli* MC4100 transformed with either pBlcEH, expressing *blc* (+), or pBlcHE, unable to express *blc* (-). Cells were induced with IPTG and subjected to French Pressure lysis. The lysates (L) were divided into soluble (S) and membrane (M) fractions by ultracentrifugation. The position of bands present during Blc expression are marked by arrowheads. Each lane corresponds to 40 μg of protein stained with Coomassie blue dye. The indicated molecular mass standards are expressed in kDa.

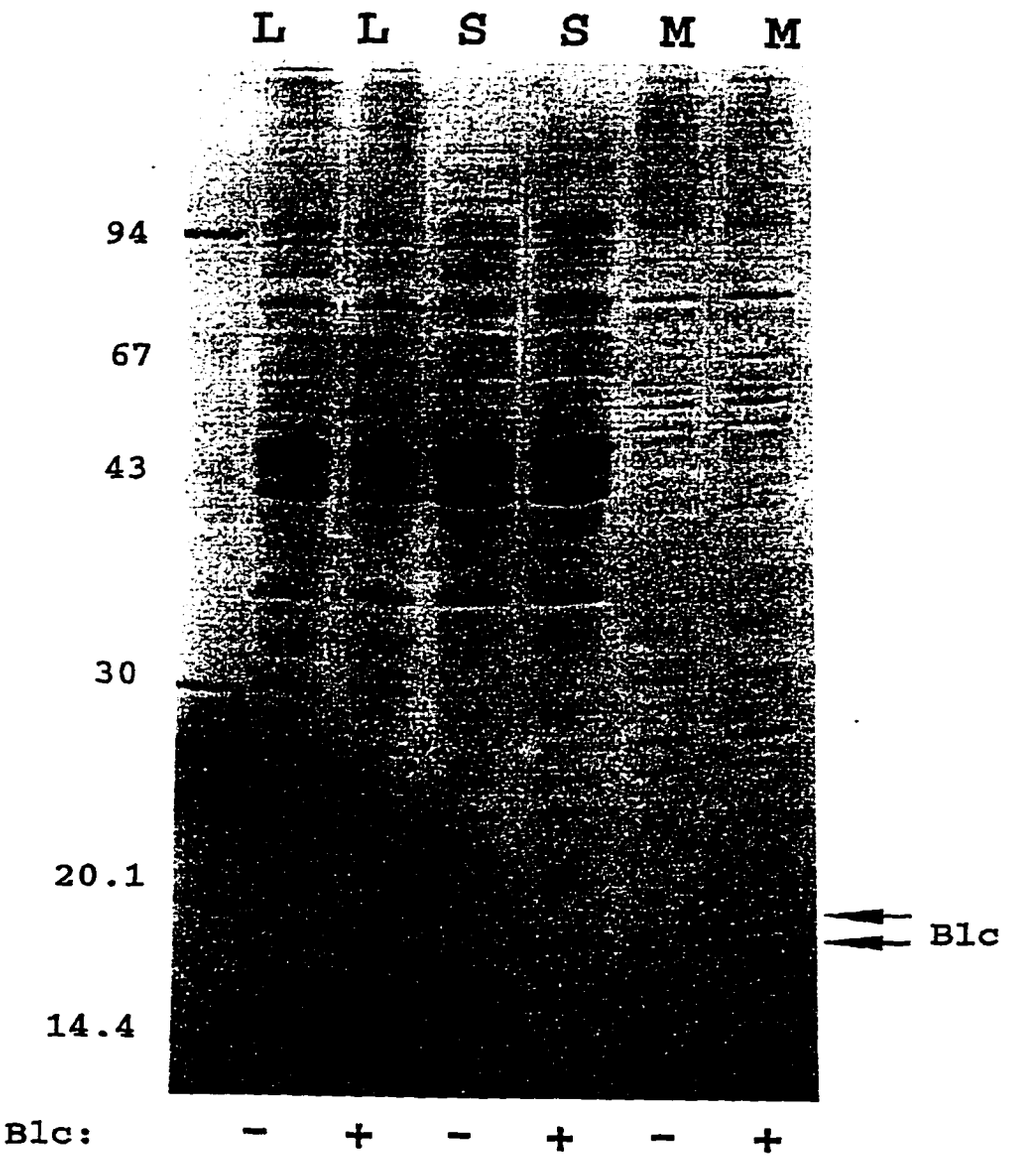

L L S S M M
94
67
43
30
20.1
14.4
Blc
Blc: - + - + - +

Figure 6-4 *Membrane fractionation of precursor and mature species of Blc*

15% SDS-PAGE analysis of *E. coli* MC4100 transformed with either pBlcEH, expressing *blc* (+), or pBlcHE, unable to express *blc* (-). Cells were induced with IPTG and subjected to French Pressure lysis. Membranes were isolated by ultracentrifugation and separated into light (L), medium (M), and heavy (H) fractions by sucrose density gradient centrifugation. The pBlcEH fractions shown at right are aligned to show the precursor and mature forms of Blc. The position of bands present during Blc expression are marked by arrowheads. Each lane corresponds to 40 μg of protein stained with Coomassie blue dye. The indicated molecular mass standards are expressed in kDa.

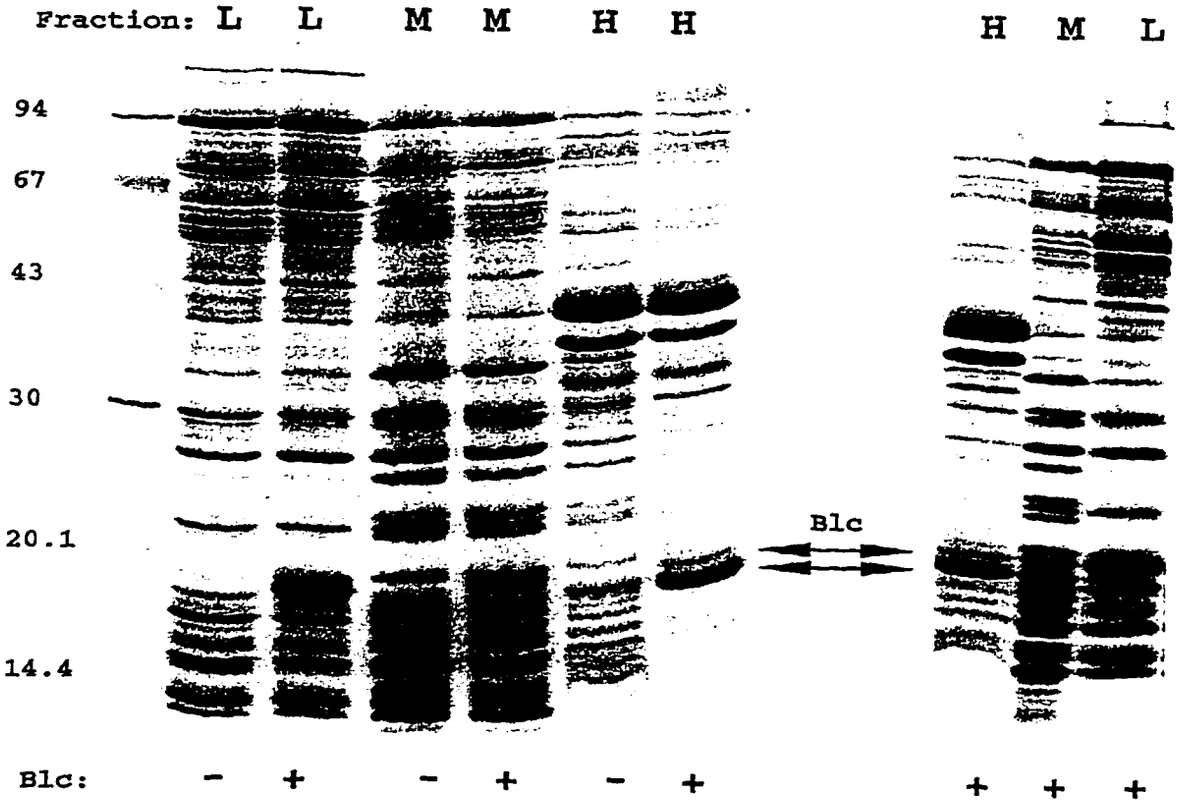

Fraction: L L M M H H H M L
94
67
43
30
20.1
14.4
Blc
Blc: - + - + - + + + +

Figure 6-5 *Metabolic labeling of Blc with [^{3}H]palmitate*

15% SDS-PAGE analysis of *E. coli* MC4100 transformed with either pBlcEH, expressing *blc* (+), or pBlcHE, unable to express *blc* (-). Cells were induced with IPTG in the presence of [^{3}H]palmitate and globomycin was added when indicated (+). Proteins were prepared by boiling the cells in SDS and samples corresponding to 800 000 cpm were resolved by 15% SDS-PAGE and visualized by fluorography. The position of bands present during Blc expression are marked by arrowheads. The indicated molecular mass standards are expressed in kDa.

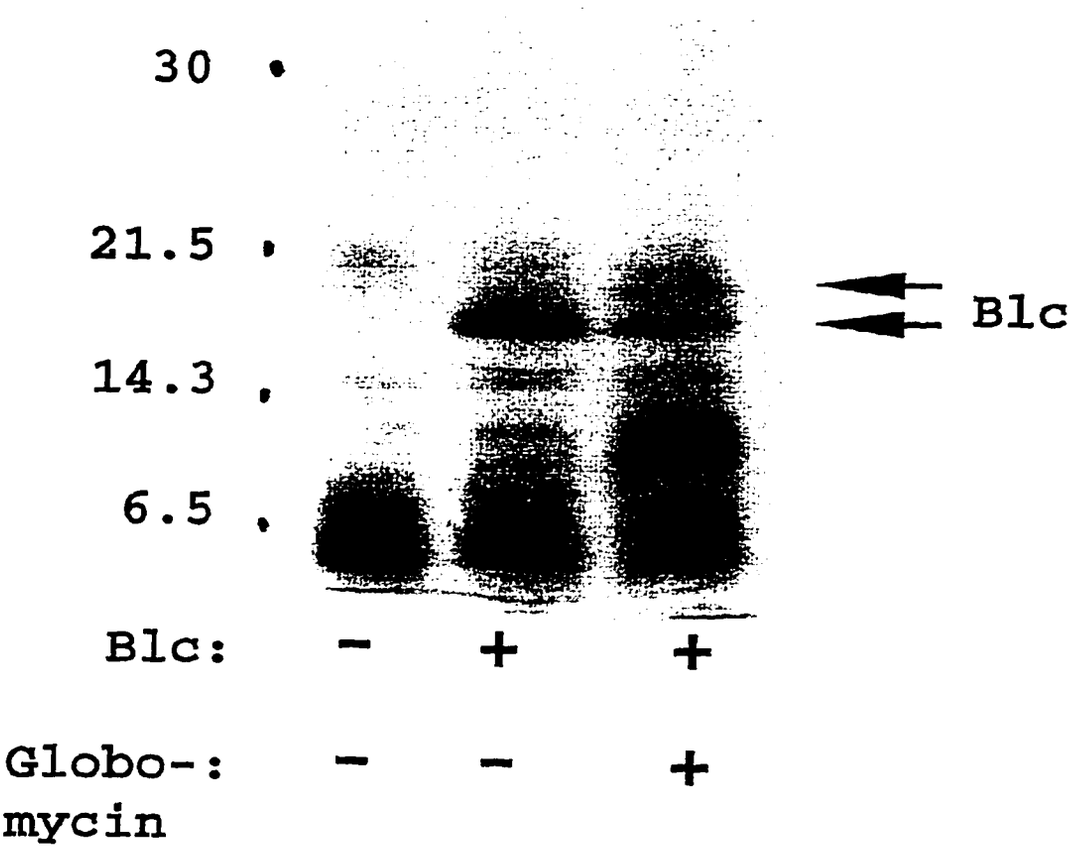

30
21.5
14.3
6.5
Blc
Blc: - + +
Globo-: - - +
mycin

TABLE 6-2

Comparison of amino acid sequence identities between Apolipoprotein-D and Bacterial Lipocalin species

The primary structures of Blc from *E. coli* (EcBlc; this paper) and *Citrobacter freundii* (CfBlc; this paper) are compared with Vlp from *Vibrio cholerae* (VcVlp; GenBank accession number x64097), Apo-D from *Mus musculus* (MmApd; GenBank accession number x82648), Apo-D from *Rattus norvegicus* (RnApd; GenBank accession number x55572), and Apo-D from *Homo sapiens* (HsApd; GenBank accession number j02611). The values represent percentage identities defined as number of identical residues between each pair of mature polypeptides as aligned in Figure 6-6, divided by the number of residues in the shortest member of the pair.

	VcVlp	EcBlc	CfBlc	MmApd	RnApd	HsApd
VcVlp	100	49	48	26	28	26
EcBlc		100	90	33	33	31
CfBlc			100	32	33	31
MmApd				100	89	74
RnApd					100	73
HsApd						100

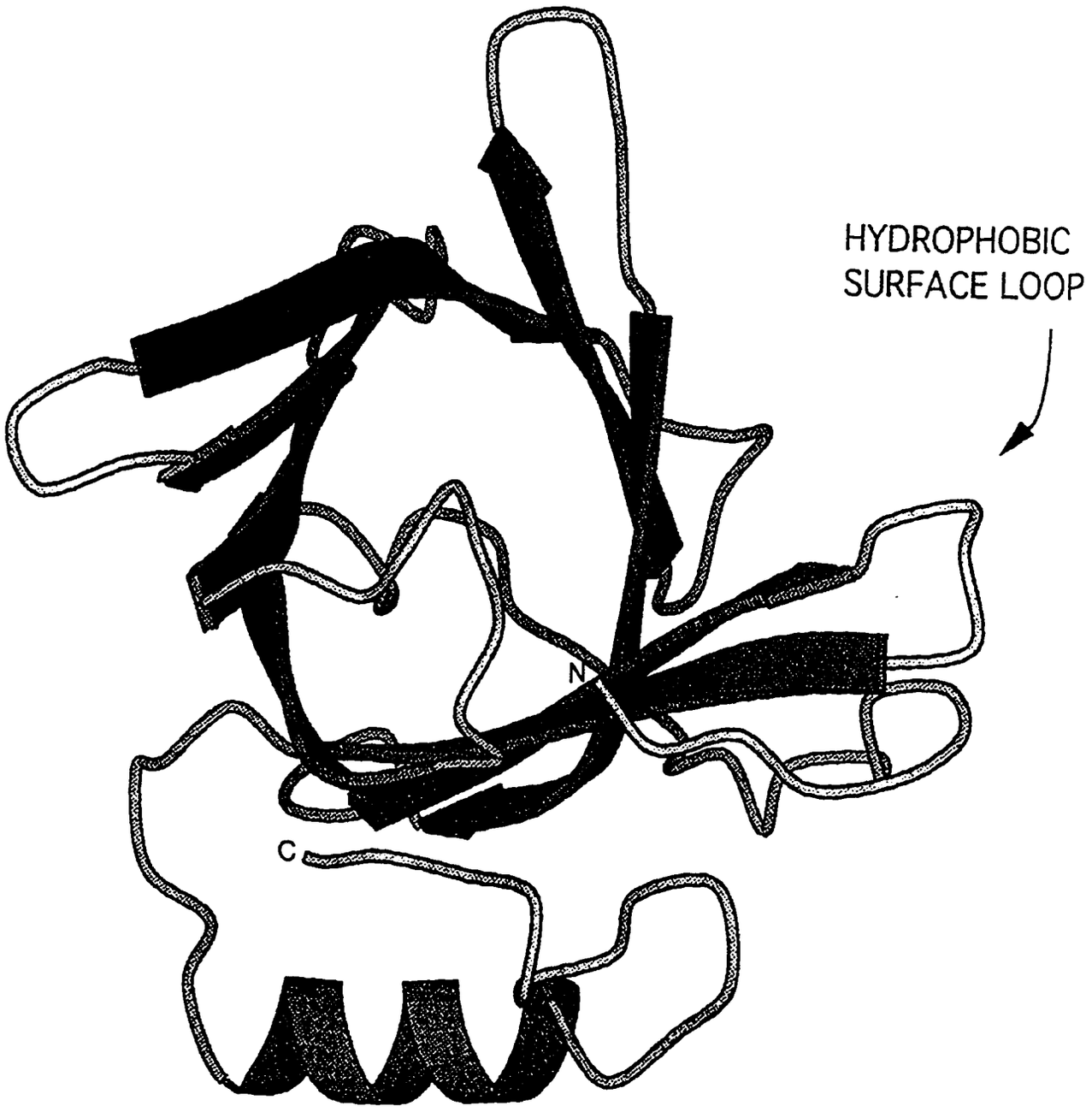
Figure 6-6 *Homologies between Apolipoprotein-D and Bacterial Lipocalins*

The Bacterial lipocalins are aligned with three Apo-D species; see Table 6-2 for the source of each sequence. Residues are shaded when at least 5 of the 6 aligned residues are identical. The positions of cysteine residues forming the two disulfide bonds in the Apo-D species are underlined in bold type and one pair distinguished by italics. Additional cysteine residues are shown in bold type and include the positions of the lipoprotein cleavage site near the amino terminus of the Vlp and Blc species; the amino terminal glutamine residues of the mature Apo-D species are also shown in bold type. The homologies were initially identified using the TFASTA algorithm (Genetics Computer Group, Wisconsin) and optimized manually by comparison with the secondary structural elements of human Apo-D, which are shown under the alignment as ß-strands (b) and α-helix (h).

```
VcVlp   --------MRAIFLILCSVLLNGCLGMPESVKPVSDFELNNYLGKWYEVA   42
EcBlc   ---MRLLPLVAAATAAFLVVACSSPTPPRGVTVVNNFDAKRYLGTWYEIA   47
CfBlc   ---MRILPVVAAVTAAFLVVACSSPTPPKGVTVVNNFDAKRYLGTWYEIA   47
MmApd   MVTMLMFLATLAGLFTTAKGQNFHLGKCPSPPVQENFDVKKYLGRWYEIE   50
RnApd   MATMLLLLATLAGLFTTTEGQSFHLGKCPSPPVQENFDVKKYLGRWYEIE   50
HsApd   MVMLLLLLSALAGLFGAAEGQAFHLGKCPNPPVQENFDVNKYLGRWYEIE   50
                                                  bbbbbbb

VcVlp   RLDHSFERGLSQVTAEYRVRNDGGISVLNRGYSEEKGEWKEAEGKAYFVN   92
EcBlc   RFDHRFERGLEKVTATYSLRDDGGLNVLNKGYNPDRGMWQQSEGKAYFTG   97
CfBlc   RFDHRFERGLDKVTATYSLRDDGGINVLNKGYNPDREMWQKTEGKAYFTG   97
MmApd   KIPASFEKG-NCIQANYSLMENGNIEVLNKELSPDG-TMNQVRGEAKQSN   98
RnApd   KIPVSFEKG-NCIQANYSLMENGNIKVLNKELRPDG-TLNQVEGEAKQSN   98
HsApd   KIPTTFENG-RCIQANYSLMENGKIKVLNQELRADG-TVNQIEGEATPVN   98
                b bbbbbbbbb      bbbbbbbb      bbbbbbbbb

VcVlp   GSTDGYLKVSFEGPF-YGSYVVFEIDRENYSYAFVSGP-------NTEYL  134
EcBlc   APTRAALKVSFEGPF-YGGYNVIALDRE-YRHALVCGP-------DRDYL  138
CfBlc   DPSTAALKVSFEGPF-YGGYNVIALDRE-YRHALVCGP-------DRDYL  138
MmApd   VSEPAKLEVQFEPLMPPAPYWIIATDYE--NYALVYSCTTFFWLFHVDFF  147
RnApd   MSEPAKLEVQFESLMPPAPYWIIATDYE--SYALVYSCTTFFWFFHVDYV  147
HsApd   LTEPAKLEVKESWFMPSAPYWIIATDYE--NYALVYSCTCIIQLFHVDFA  147
                bbbbbbb      bbbbbbbbb      bbbbbbbbb

VcVlp   WILSRTPTVERGILDKFIEMSKERGFDTNRLIYVQLQ------------  171
EcBlc   WILSRTPTISDEVKQEMLAVATREGFDVSKFIWVQQPGS----------  177
CfBlc   WILSRTPTISDEMKQQMLAIATREGFEVNKLIWVKQPGA----------  177
MmApd   WILGRNPYLPPETITYYLKDILTSNGIDIEKMTTTDQANCPDFL      189
RnApd   WILGRNPYLPPETITYYLKYILTSNDIDIAKITTKDQANCPDFL      189
HsApd   WILARNPNLPPETVDSLKNILTSNNIDVKKMTVTDQVNCPKLS      189
        bbb         hhhhhhhhhhh
```

Figure 6-7 *Molecular model of Human Apolipoprotein-D*

Molecular model of Human Apo-D emphasizing the 8 antiparallel ß-strands, which form two orthogonal sheets, followed by a C-terminal α-helix. The hydrophobic surface loop that separates ß-strands 7 and 8 is marked with an arrowhead. This MOLSCRIPT diagram was generated by Natalie Strynadka using the coordinates of human Apo-D obtained from the Brookhaven Protein Data Bank under accession number 2APD.


HYDROPHOBIC
SURFACE LOOP
N
C

REFERENCES

(1) Flower, D. R., North, A. C. T., and Attwood, T. K. (1993) *Protein Sci.* **2,** 753-761

(2) Newcomer, M. E., Jones, T. A., Åqvist, J., Sundelin, J., Eriksson, U., Rask, L., and Peterson, P. A. (1984) *EMBO J.* **3,** 1451-1454

(3) Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., and Kayser, H. (1987) *J. Mol. Biol.* **198,** 499-513

(4) Holden, H. M., Rypniewski, W. R., Law, J. H., and Rayment, I. (1987) *EMBO J.* **6,** 1565-1570

(5) Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E., and Kraulis, P. J. (1986) *Nature* **324,** 383-385

(6) Godovac-Zimmermann, J. (1988) *TIBS* **13,** 64-66

(7) Drayna, D., Fielding, C., McLean, J., Baer, B., Castro, G., Chen, E., Comstock, L., Henzel, W., Kohr, W., Rhee, L., Wion, K., and Lawn, R. (1986) *J. Biol. Chem.* **261,** 16535-16539

(8) Peitsch, M. C., and Boguski, M. S. (1990) *New Biol.* **2,** 197-206

(9) Balbin, M., Freije, J. M. P., Fueyo, A., Sánchez, L. M., and López-Otín, C. (1990) *Biochem. J.* **271,** 803-807

(10) Pearlman, W. H., Gueriguian, J. L., and Sawyer, M. E. (1973) *J. Biol. Chem.* **248,** 5736-5741

(11) López-Boado, Y. S., Tolivia, J., and López-Otín, C. (1994) *J. Biol. Chem.* **269,** 26871-26878

(12) Blais, Y., Sugimoto, K., Carriere, M. C., Haagensen, D. E., Labrie, F., and Simard, J. (1994) *Int. J. Cancer* **59,** 400-407

(13) Boyles, J. K., Notterpek, L. M., and Anderson, L. J. (1990) *J. Biol. Chem.* **265,** 17805-17815

(14) Spreyer, P., Schaal, H., Kuhn, G., Rothe, T., Unterbeck, A., Olek, K., and Müller, H. W. (1990) *EMBO J.* **9,** 2479-2484

(15) Sansom, C. E., North, A. C. T., and Sawyer, L. (1994) *Biochim. Biophys. Acta* **1208,** 247-255

(16) Braun, V., and Wu, H. C. (1993) in *Bacterial Cell Wall, New Comprehensive Biochemistry* (Ghuysen, J. -M., and Hackenbeck, R., eds.) Vol. 27, pp. 319-342, Elsevier Science Publishers B. V., Amsterdam

(17) Westenberg, D. J., Gunsalus, R. P., Ackrell, B. A. C., Sices, H., and Cecchini, G. (1993) *J. Biol. Chem.* **268,** 815-822

(18) Lange, R., and Hengge-Aronis, R. (1991) *J. Bacteriol.* **173,** 4474-4481

(19) Miller, J. H. (1992) *A Short Course in Bacterial Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

(20) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

(21) Lowry, O. H., Rosenbrough, N. J., Farr, A. L., and Randall, R. J. (1951) *J. Biol. Chem.* **193,** 265-275

(22) Bishop, R. E., and Weiner, J. H. (1993) *Eur. J. Biochem.* **213,** 405-412

(23) Clarke, L. and Carbon, J. (1976) *Cell* **9:** 91-99

(24) Jaurin, B., and Grundström, T. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4897-4901

(25) Greener, T., Govezensky, D., and Zamir, A. (1993) *EMBO J.* **12,** 889-896

(26) Frost, L., Lee, S., Yanchar, N., and Paranchych, W. (1989) *Mol. Gen. Genet.* **218,** 152-160

(27) Jain, C.(1993) *Gene* **133,** 99-102

(28) Mulvey, M.R., and Loewen, P.C. (1989) *Nucleic Acids Res.* **17,** 9979-9991

(29) Youngman, P. (1987) in *Plasmids: a practical approach* (Hardy, K.G., ed.) pp. 79-102, IRL Press, Oxford, U.K.

(30) Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992) *J. Biol. Chem.* **267,** 13062-13072

(31) Walderich, B., Ursinus-Wössner, A., van Duin, J., and Höltje, J.-V. (1988) *J. Bacteriol.* **170,** 5027-5033

(32) Jones, T.A. (1985) *Methods Enzymol.* **115,** 157-171

(33) Kraulis, P.J. (1991) *J. Appl. Crystallogr.* **24,** 946-950

(34) Bishop, R. E., and Weiner, J. H. (1993) *FEMS Microbiol. Lett.* **114,** 349-354

(35) Condon, C., and Weiner, J. H. (1988) *Molec. Microbiol.* **2,** 43-52

(36) Kaasen, I., Falkenberg, P., Stryvold, O. B, and Strøm, A. R. (1992) *J. Bacteriol.* **174,** 889-898

(37) Hengge-Aronis, R. (1993) *Cell* **72,** 165-168

(38) Tanaka, K, Takayanagi, Y, Fujita, N., Ishihama, A., and Takahashi, H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3511-3515

(39) Nguyen, L. H., Jensen, D. B., Thompson, N. E., Gentry, D. R., and Burgess, R. R. (1993) *Biochemistry* **32,** 11112-11117

(40) Hussain, M., Ichihara, S., and Mizushima, S. (1980) *J Biol Chem* **255,** 3707-3712

(41) Groat, R. G., Schultz, J. E., Zychlinsky, E., Bockman, A., and Matin, A. (1986) *J. Bacteriol.* **168,** 486-493

(42) Matin, A., Augen, G. A., Blum, P. H., and Schultz, J. E. (1989) *Ann. Rev. Microbiol.* **43,** 293-316

(43) Alexander, D.M., and St John, A.C. (1994) *Molec. Microbiol.* **11,** 1059-1071

(44) Hengge-Aronis, R., Klein, W., Lange, R., Rimmele, M., and Boos, W. (1991) *J. Bacteriol.* **173,** 7918-7924

(45) Yang, C. -Y., Gu, Z. -W., Blanco-Vaca, F., Gaskell, S. J., Yang, M., Massey, J. B., Gotto, A. M. Jr., and Pownall, H. J. (1994) *Biochemistry* **33,** 12451-12455

(46) Yamguchi, K., Yu, F., and Inouye, M. (1988) *Cell* **53,** 423-432

CHAPTER 7

Bacterial Lipoproteins as Intermediates in Membrane Protein Evolution[1]

Energy is Eternal Delight.

William Blake

[1]A portion of this chapter has been published. Bishop, R.E., and Weiner, J.H. (1996) *Trends Biochem. Sci.* **21:** 127

Introduction

We have previously identified a bacterial ancestor of the lipocalin family of lipid-binding proteins (1). The lipocalins were once thought to be restricted to Eukarya, but they clearly evolved from a bacterial outer membrane lipoprotein. The mechanism of membrane association involves the modification of an amino terminal cysteine residue, which is specified by a type II signal peptide. The cysteine is modified by a thioether-linked diglyceride prior to cleavage of the signal peptide by signal peptidase II. The nascent α-amino group is then modified in enterobacteria by an amide linked fatty acid. The resulting lipoprotein is thereby anchored to the bacterial cell envelope. We have found that not only lipocalins, but also cupredoxins, have been adapted through evolution to interact with membranes as lipoproteins. While the lipocalins appear to have originated as peripheral membrane proteins that were later released into the aqueous milieu through the secondary loss of the lipoprotein anchor, the cupredoxins appear to have acquired the lipoprotein anchor prior to becoming integrated into membranes as subunit II of cytochrome oxidase.

Methods

Computer methods, lipoprotein modification, ß-galactosidase analysis, and bacterial growth were performed as described (1). For chromophore analysis, 25mL cultures of *E. coli* MC4100/pBlcEH cells were grown for 4 hours prior to induction with IPTG for 6-8 hrs, harvested, and washed as described (1). Cells were centrifuged in 1mL Eppendorf tubes and extracted with 1/2 vol. of chloroform, which was pooled and analyzed in quartz cuvettes using a Varian DMS 100S UV visible spectrophotometer. Plasmids pRC07 and pC25A in the *cyo-cyd* double mutant *E.coli* GO105 (provided by Dr. R.B. Gennis[26]) were grown on the medium described (1) supplemented with 0.3% D,L-lactate to stimulate aerobic respiration; plasmid pRC07 carries the wild-type *cyo* operon and pC25A is an isogenic derivative with a Cys25Ala mutation in *cyoA*.

Results and Discussion

"Outlier" Lipocalins more than Peripheral

The lipocalins are a diverse and burgeoning family of primarily extracellular proteins involved in the binding and transport of small hydrophobic ligands. Until recently, lipocalins were restricted to eukaryotes, but the appearance of three bacterial lipocalins (Blc) suggests that they originated earlier in evolution than previously imagined (1). In the absence of an Archaeal lipocalin, it is unknown whether the Eukarya obtained the lipocalins by horizontal or vertical transfer. Nevertheless, the Blc proteins are grouped with "outliers" from the larger cluster of "kernel" lipocalins due to an absence of intramolecular disulfide bonds (2). They are most closely related to a subgroup of lipocalins also distinguished by their disulfide bonding pattern; these include mammalian apolipoprotein D (ApoD) and insect Lazarillo (Laz). Two enterobacterial Blc proteins share 34-39% amino acid sequence identity with four mammalian ApoD's, whereas a Blc relative from *Vibrio cholerae* shares 31% identity with Laz. Comparisons of enterobacterial Blc with Laz and *Vibrio* Blc with ApoD yield lower values, perhaps indicating two distinct lines of orthologous descent, and also reflecting the intrinsically low primary structural similarity between lipocalins (3). But the similarities only begin at the amino acid sequence level (Fig. 7-1).

What distinguishes Blc, ApoD, and Laz from the other lipocalins biochemically is their peripheral membrane localization. However, they associate with membranes by distinct mechanisms: Blc is anchored to the outer membrane of the Gram-negative cell envelope by a lipid-modified amino-terminal cysteine residue (ie. Blc is a typical prokaryotic lipoprotein [1]), ApoD associates with plasma lipoproteins by a solvent exposed hydrophobic loop on the protein surface (4,5), and Laz is anchored to membranes by a glycosyl-phosphatidylinositol (GPI) anchor at the carboxy-terminus (6).

The diverse membrane anchor mechanisms within this group of lipocalins may reflect the adaptation of eukaryotic cells to the absence of corresponding prokaryotic lipoprotein modification enzymes. It appears that the simplest solution was to insert 5 or 6 hydrophobic residues into a solvent exposed surface loop to ensure the correct membrane localization (Fig. 7-2); if this adaptation has remained with ApoD, it seems that Laz followed with the acquisition of a carboxy-terminal signal sequence necessary for the highly specialized GPI anchor. Once Laz acquired the GPI anchor, the residues in the hydrophobic surface loop were probably converted to more appropriate hydrophilic residues. This model is consistent with the same insertion, between ß-strands 7 and 8, and the same pattern of disulfide bonds in both Laz and ApoD (Figs. 7-1 and 7-2).

The common membrane localization of Blc, ApoD, and Laz may provide clues regarding the evolution of lipocalins as soluble proteins designed to transport hydrophobic ligands. The removal of membrane anchors, either enzymatically or genetically, would liberate the lipocalin into the aqueous milieu. The most closely related soluble lipocalins are the insecticyanins, which bind a biliverdin chromophore, and provide crystal structure coordinates (7,8) for molecular modeling studies. Such information will provide useful guidelines in the search for a common ligand that may be shared among the membrane anchored lipocalins. We find that overproduction of Blc results in the accumulation of a blue chromophore, which associates with the cell envelope and is chloroform extractable. The absorption spectrum (Fig. 7-3) exhibits a maximum near 605 nm with a shoulder at 570 nm; these features are shared with the chlorin of heme D (9), but we were unable to confirm the identification because the Soret band is obscured in our crude preparation. Purification and chemical analysis of the blue chromophore associated with Blc overproduction will be required in order to determine whether there is any physical association between the two entities.

Additionally, both the *blc* and *apoD* genes are activated in response to starvation conditions; *blc* is controlled by the RpoS stationary phase sigma factor (1), and *apoD* is

expressed in non-proliferating quiescent and senescent fibroblasts (10), suggesting that both genes fulfill similar physiological functions. We have discovered that the *blc* promoter belongs to the subclass of RpoS-dependent promoters that depend on dual regulation by starvation and osmotic stress (11). LacZ activity governed by the *blc* promoter is expressed optimally in stationary phase when 300mM NaCl is included in the growth medium (Fig. 7-4); the activity is abolished when bacteria are grown in the absence of salt. Our earlier measurements (1) gave intermediate values because we grew the bacteria under standard salt concentrations of 171 mM.

It was recently demonstrated that the nervous system is a primary source of *apoD* transcription (12), and Laz expression is restricted primarily to a subset of insect neural cells (13). Again, the common neural expression of ApoD and Laz suggests that these proteins have a related function: could certain aspects of this function also be exerted in the bacterial cell envelope? These common features provide a fascinating perspective on this newly uncovered subgroup of lipocalins. Perhaps the lipocalins anchored at the membrane periphery are less likely "outliers" and more likely the primordial ancestors of the entire family.

Lipidation of the Blue Copper Proteins

Another class of blue proteins that evolved from lipoprotein ancestors are the cupredoxins. In this case, it appears that evolution started with a soluble ancestor that was later adapted as a lipoprotein. The cupredoxins belong to a large family of copper containing proteins that evolved from a common ancestor (14). The cupredoxins possess a Greek Key ß-barrel structure and are found in Archaea, Bacteria, and Eukarya. A representative group, the chloroplast plastocyanins, contain a single copper atom and participate in the photosynthetic electron transfer chain. Similarly, the azurins are involved in electron transfer to cytochrome oxidase in the bacterial respiratory chain. Azurins of *Neisseria* have been identified that possess a type II signal peptide for

lipoprotein processing (15,16) (Table 7-I). The cupredoxin of *N. pharaonis*, called halocyanin, is the only known Archaeal lipoprotein (17) and demonstrates that Archaea, despite their unique lipids, also utilize type II signal peptidase machinery, which, like the cupredoxins, must have existed in the last common ancestor.

The adaptation of lipoprotein cupredoxins for electron transfer reactions at the periphery of energy transducing membranes may have provided a selective advantage by restricting the protein to lateral diffusion. This appears to have been followed by the incorporation of a cupredoxin-like subunit into the integral membrane complex of cytochrome oxidase (18). The binuclear copper containing subunit II of cytochrome oxidase is derived from the cupredoxins (19); some subunits II have lost the binuclear copper centers and function instead as quinol oxidases. Subunits II of cytochrome and quinol oxidases are anchored to membranes by a pair of amino terminal transmembrane α-helical domains. The quinol oxidase subunit II of *E. coli* known as CyoA was found to possesses a typical type II signal peptide during an inspection of the upstream operon that encodes AmpG (20). In fact, we found that a group of subunits II contain the same lipoprotein signature (Table 7-I). By utilizing a Cys25Ala mutant of CyoA expressed in a *cyo-cyd*-deletion strain provided by Dr. R.B. Gennis, wild-type CyoA was shown to incorporate [^{3}H]-palmitate (Fig. 7-5). The mutant is unable to grow by aerobic respiration unless it is transformed with a plasmid bearing a functional *cyo* operon. While the strain bearing the plasmid with the Cys25Ala mutation in *cyoA* could grow aerobically (Fig. 7-6), it exhibited a lag in the onset of growth when colonies from plates were inoculated into liquid medium. We suggest that the block in lipoprotein modification of CyoA delays the assembly of active cytochrome oxidase, but does not prevent the function of the enzyme once assembled.

The type II signal peptide associated with subunits II of cytochrome oxidase is probably an evolutionary vestige of a primordial cupredoxin lipoprotein that was incorporated into the ancestral subunit. The type II signal peptide is present among both

cytochrome and quinol oxidases in Gram positive bacteria, but has been lost from the cytochrome oxidase transferred vertically to the proteobacteria and, subsequently, to mitochondria. This adaptation may have been essential since the Eukarya do not possess type II signal peptidase machinery. However, the quinol oxidase was transferred to the proteobacteria by a later horizontal gene transfer event (21) and still contains the type II signal peptide as represented by CyoA. The susceptibility of the lipoprotein subunits II of Gram positive bacteria to Edman degradation (22-24) suggests that they do not possess the amide linked fatty acid, which is also absent in the lipoprotein cytochrome subunit of the Photosynthetic Reaction Center (25). However, the N-terminus of CyoA is blocked as expected of enterobacterial lipoproteins (26). The discovery of Archaeal subunits II with single transmembrane α-helical segments provides an additional variation (27) that is incorporated into an evolutionary scheme (Fig 7-7). Incredibly, these aerobic respiratory enzymes originated before oxygen was released into the atmosphere (28).

Figure 7-1:

Multiple sequence alignments performed using the Genetics Computer Group algorithm PILEUP. Residues involved in membrane anchoring are enboldened and marked with an asterisk. Cysteine residues involved in the formation of intramolecular disulfide bonds are enboldened and underlined, and one pair distinguished by italics. The secondary structural elements derived from the molecular model of human ApoD (pdb code 2APD), which was based on the coordinates of insecticyanin (4), are listed below the sequence as beta-sheet (bbb) or alpha-helix (hhh). Identical residues are shaded and conserved residues outlined. Sequences of the bacterial lipocalins (Blc), lazarillo (Laz), and apolipoproteins D (Apd), are given species designations and GenBank accession codes: *Escherichia coli* (Ec; U21726), *Citrobacter freundii* (Cf; U21727), *Vibrio cholerae* (Vc; X64097), *Schistocerca americana* (Sa; Z38071), *Mus musculus* (Mm; L39123), *Rattus norvegicus* (Rn; X55572), *Oryctolagus cuniculus* (Oc; L42979), and *Homo sapiens* (Hs; J02611).

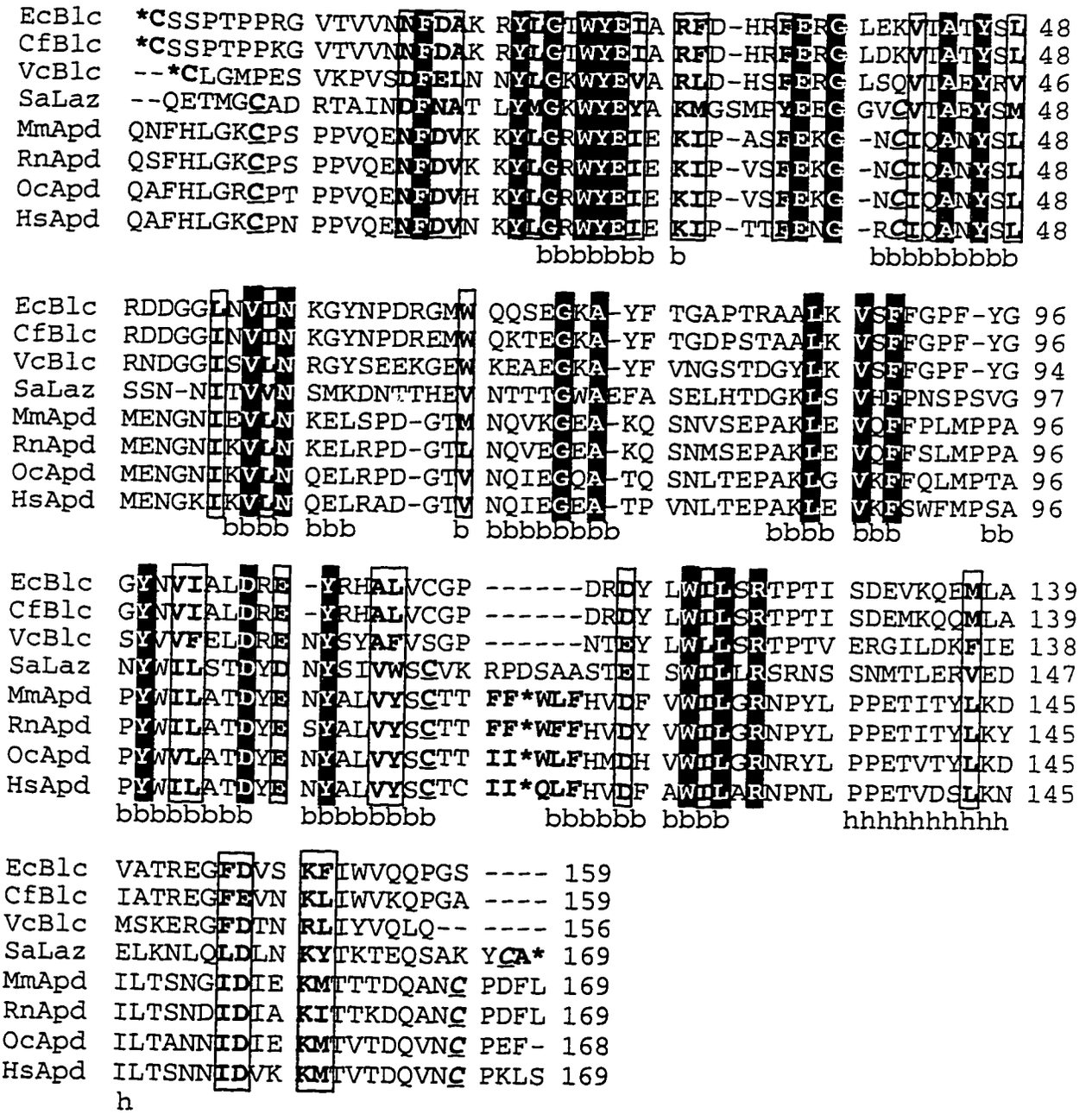

EcBlc *CSSPTPPRG VTVVNNFDAK RYLGIWYEIA RFD-HRFERG LEKVIATYSL 48
CfBlc *CSSPTPPKG VTVVNNFDAK RYLGIWYEIA RFD-HRFERG LDKVIATYSL 48
VcBlc --*CLGMPES VKPVSDFELN NYLGKWYEVA RLD-HSFERG LSQVIAEYRV 46
SaLaz --QETMGCAD RTAINDFNAT LYMGKWYEYA KMGSMPYEEG GVCVTAEYSM 48
MmApd QNFHLGKCPS PPVQENFDVK KYLGRWYEIE KIP-ASFEKG -NCIQANYSL 48
RnApd QSFHLGKCPS PPVQENFDVK KYLGRWYEIE KIP-VSFEKG -NCIQANYSL 48
OcApd QAFHLGRCPT PPVQENFDVH KYLGRWYEIE KIP-VSFEKG -NCIQANYSL 48
HsApd QAFHLGKCPN PPVQENFDVN KYLGRWYEIE KIP-TTFENG -RCIQANYSL 48
bbbbbbb b bbbbbbbbb
EcBlc RDDGGLNVIN KGYNPDRGMW QQSEGKA-YF TGAPTRAALK VSFFGPF-YG 96
CfBlc RDDGGINVIN KGYNPDREMW QKTEGKA-YF TGDPSTAALK VSFFGPF-YG 96
VcBlc RNDGGISVLN RGYSEEKGEW KEAEGKA-YF VNGSTDGYLK VSFFGPF-YG 94
SaLaz SSN-NIIVVN SMKDNTTHEV NTTTGWAEFA SELHTDGKLS VHFPNSPSVG 97
MmApd MENGNIEVLN KELSPD-GIM NQVKGEA-KQ SNVSEPAKLE VQFFPLMPPA 96
RnApd MENGNIKVLN KELRPD-GTL NQVEGEA-KQ SNMSEPAKLE VQFFSLMPPA 96
OcApd MENGNIKVLN QELRPD-GTV NQIEGQA-TQ SNLTEPAKLG VKFFQLMPTA 96
HsApd MENGKIKVLN QELRAD-GTV NQIEGEA-TP VNLTEPAKLE VKFSWFMPSA 96
bbbb bbb b bbbbbbbbb bbbb bbb bb
EcBlc GYNVIALDRE -YRHALVCGP ------DRDY LWILSRTPTI SDEVKQEMLA 139
CfBlc GYNVIALDRE -YRHALVCGP ------DRDY LWILSRTPTI SDEMKQQMLA 139
VcBlc SYVVFELDRE NYSYAFVSGP ------NTEY LWLLSRTPTV ERGILDKFIE 138
SaLaz NYWILSTDYD NYSIVWSCVK RPDSAASTEI SWILLRSRNS SNMTLERVED 147
MmApd PYWILATDYE NYALVYSCTT FF*WLFHVDF VWILGRNPYL PPETITYLKD 145
RnApd PYWILATDYE SYALVYSCTT FF*WFFHVDY VWILGRNPYL PPETITYLKY 145
OcApd PYWVLATDYE NYALVYSCTT II*WLFHMDH VWILGRNRYL PPETVTYLKD 145
HsApd PYWILATDYE NYALVYSCTC II*QLFHVDF AWILARNPNL PPETVDSLKN 145
bbbbbbbb bbbbbbbb bbbbbb bbbb hhhhhhhhhh
EcBlc VATREGFDVS KFIWVQQPGS ---- 159
CfBlc IATREGFEVN KLIWVKQPGA ---- 159
VcBlc MSKERGFDTN RLIYVQLQ-- ---- 156
SaLaz ELKNLQLDLN KYTKTEQSAK YCA* 169
MmApd ILTSNGIDIE KMTTTDQANC PDFL 169
RnApd ILTSNDIDIA KITTKDQANC PDFL 169
OcApd ILTANNIDIE KMTVTDQVNC PEF- 168
HsApd ILTSNNIDVK KMTVTDQVNC PKLS 169
h

Figure 7-2:

Cartoon depicting a hypothetical scheme for the observed phylogenetic distribution of the membrane associated lipocalins. The characteristic lipocalin fold, comprises eight antiparallel ß-strands, which form two stacked orthogonal sheets (depicted as arrows), followed by a C-terminal α-helix (depicted as a barrel) (3). The diacylglyceryl moieties, associated with the N-terminal modification of Blc and the C-terminal GPI-anchor of Laz, are drawn schematically as perpendicular rectangles. The hydrophobic surface loop between ß-strands 7 and 8 is also indicated as a smaller open rectangle.

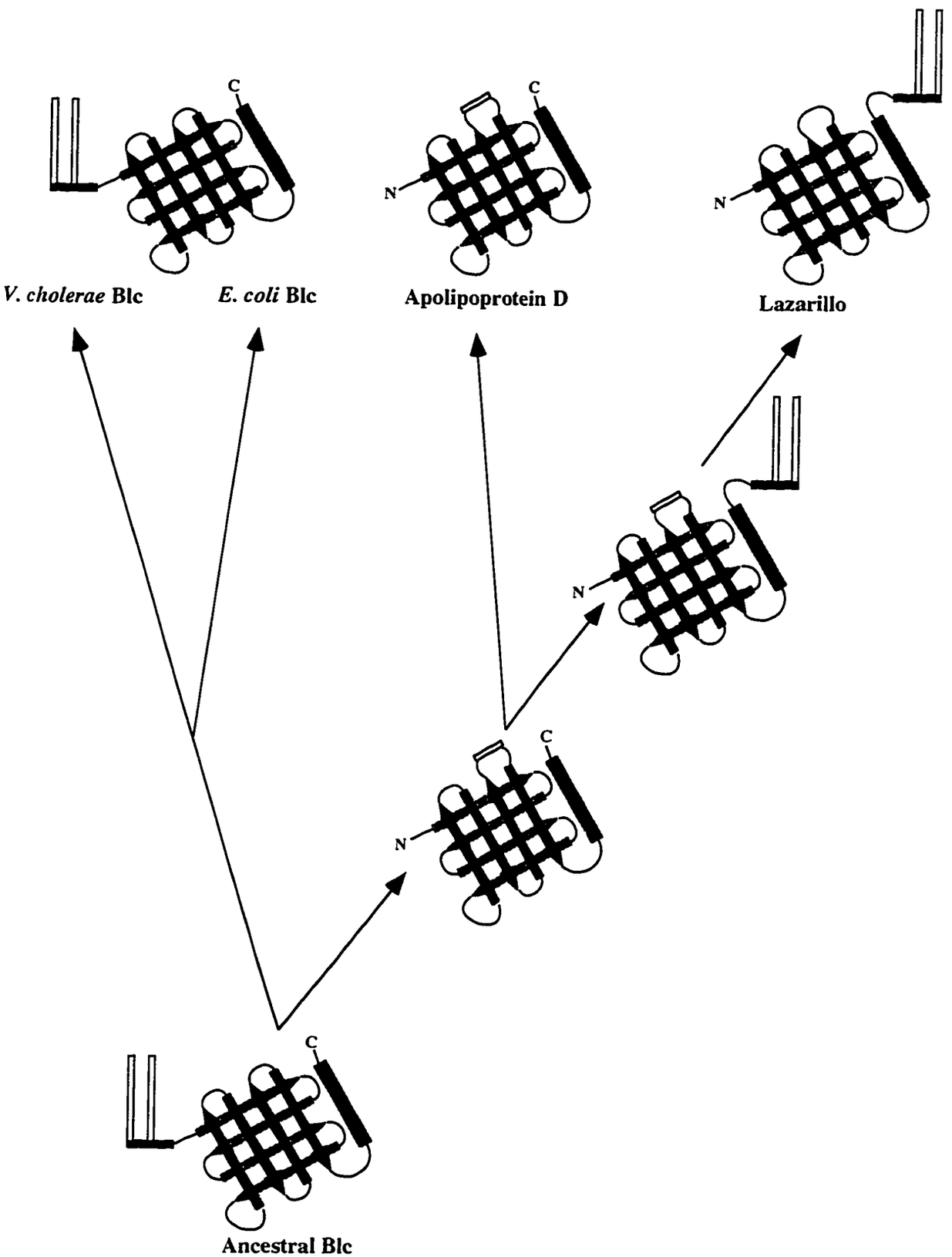
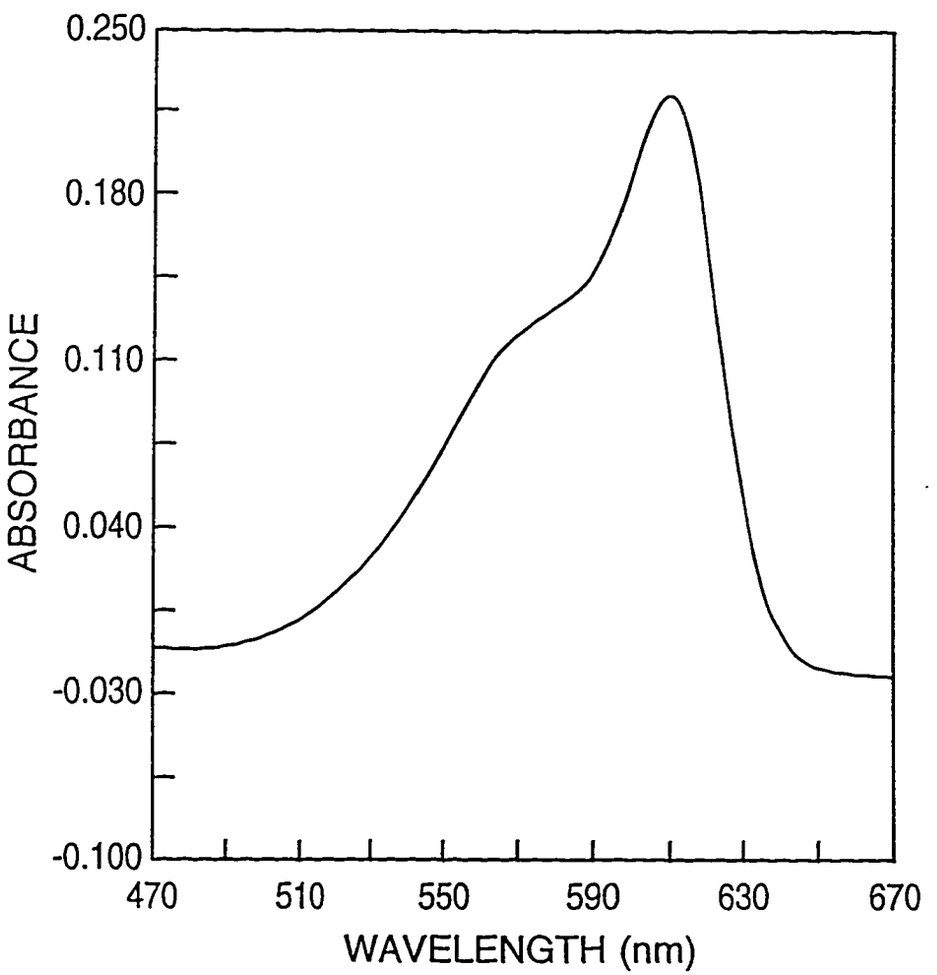

C
N
V. cholerae Blc
E. coli Blc
Apolipoprotein D
Lazarillo
Ancestral Blc

Figure 7-3:

Visible absorption spectrum of chloroform extracts from *E. coli* MC4100/pBlcEH cells producing a blue chromophore associated with Blc overexpression by IPTG induction (1).


0.250
0.180
0.110
0.040
-0.030
-0.100
ABSORBANCE
470
510
550
590
630
670
WAVELENGTH (nm)

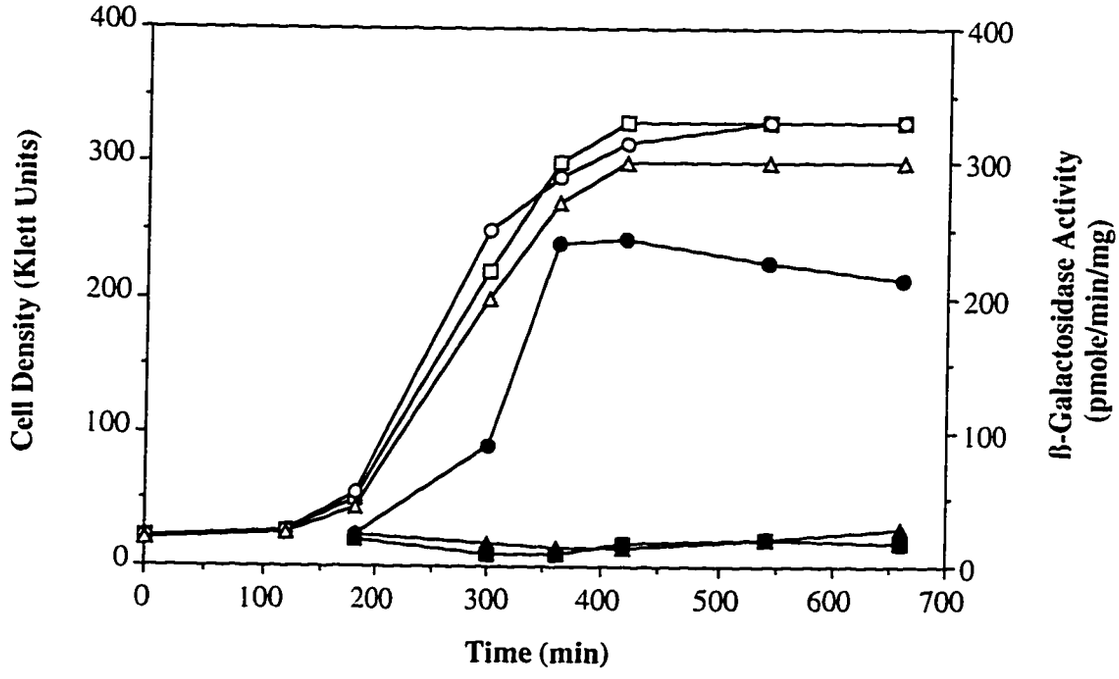


Figure 7-4 Growth of *E. coli* MC4100 (wild-type) in the absence (□) or presence of 300mM NaCl (○) and an *rpoS359::Tn*10 mutant (*E. coli* RH90) in the presence of 300mM NaCl (△). Solid symbols corresponding to the growth curves indicate ß-galactosidase expressed from the *blc::lacZ* translational fusion plasmid pBlcLacEK (1), which was carried by the strains.

Table 7-1. Putative Lipoprotein Signal Peptides of Cupredoxin Homologs

GenBank		+1	Cu[1]	Organism[2]
j05492	MRLRKYNKSLGWLSLFAGTVLLSG	CN	-	*Escherichia coli*
d13185	MKNKLLARVARLGGLSSALLLAG	CE	-	*Acetobacter aceti*
x78196	MTYIRKFARLPWLALLIPLAA	CK	-	*Paracoccus denitrificans*
m86548	MVIFLFRALKPLLVLALLTVVFVLGG	CS	-	*Bacillus subtilis*
x73124	MIFLFRALKPLLVLALLTVVFVLGG	CS	-	*Bacillus subtilis*
d13955	MNKGLCNWRLFSLFGMMALLLAG	CG	II	*Bacillus PS3*
m94110	MKLWKTASRFLPLSFLTLFLTG	CL	II	*Bacillus firmus*
d85547	MKRWLPNLRVFSLFSVLALFLAG	CG	II	*Bacillus sp.*
d70843	MKKGLRNWRLFSLFGMMALLLAG	CG	II	*Bacillus stearothermophilus*
x54140	MVKHWRLILLLALVPLLLSG	CG	II	*Bacillus subtilis*
z30326	MKDISRRRFVLGTGATVAAATLAG	CN	I	*Natronobacterium pharaonis*
x06208	MKAYLALISAAVIGLAA	CS	I	*Neisseria gonorrhoeae*
y00530	MKAYLALISAAVIGLAA	CS	I	*Neisseria meningitidis*

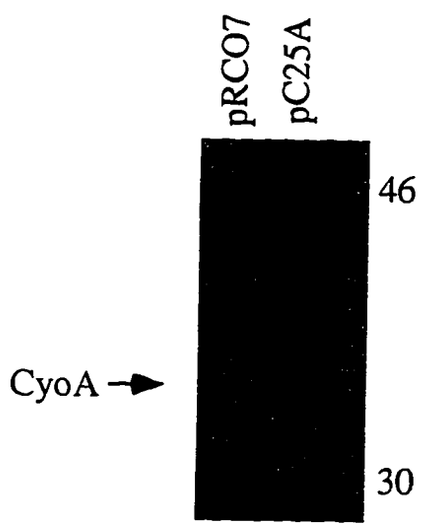
1. The quinol oxidases have lost the type II copper that remains with subunit II of the cytochrome oxidases, while type I copper is associated with cupredoxins.

2. *Bacillus spp.* are Gram positive, *Natronobacterium* is an Archaeon, and the remainder are proteobacteria. Conserved residues characteristic of lipoprotein signal peptides are shown in bold. The signal peptidase II cleavage site is indicated by +1.

m86548: 11 residues truncated from author-given translational initiation site

Figure 7-5:

12% SDS-PAGE analysis of *E. coli* GO105 (*cyd, cyo*) carrying the wild type *cyo* operon on plasmid pRC07 or pC25A, which carries a Cys25Ala mutation in the *cyoA* cistron. Cells were grown in LB medium supplemented with 0.3% D,L-lactate in the presence of [^{3}H]-palmitate and analyzed by fluorography as described (1).


pRCO7
pC25A
46
CyoA
30

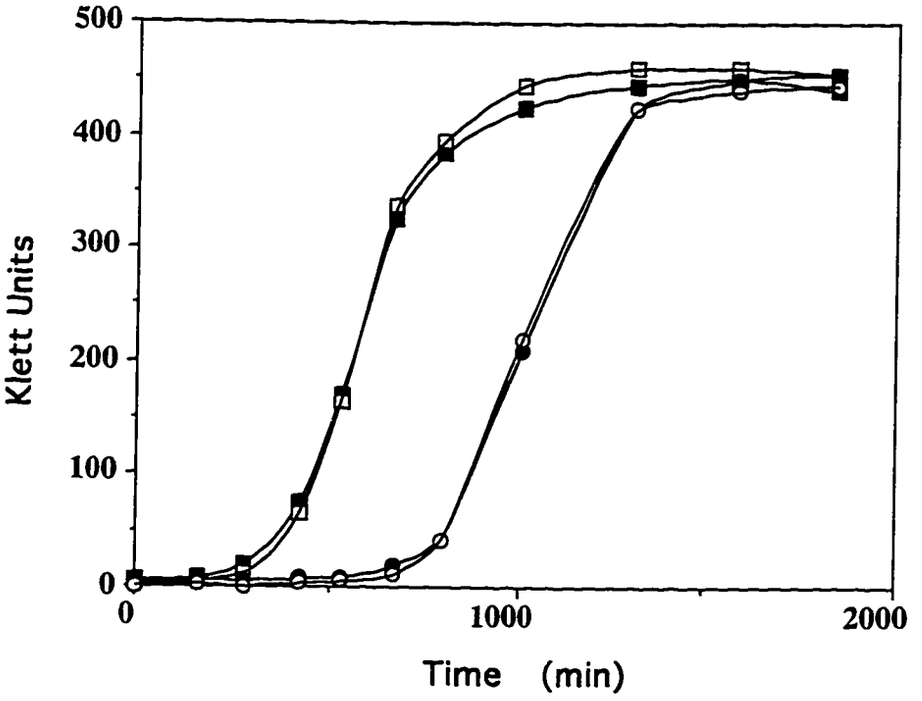
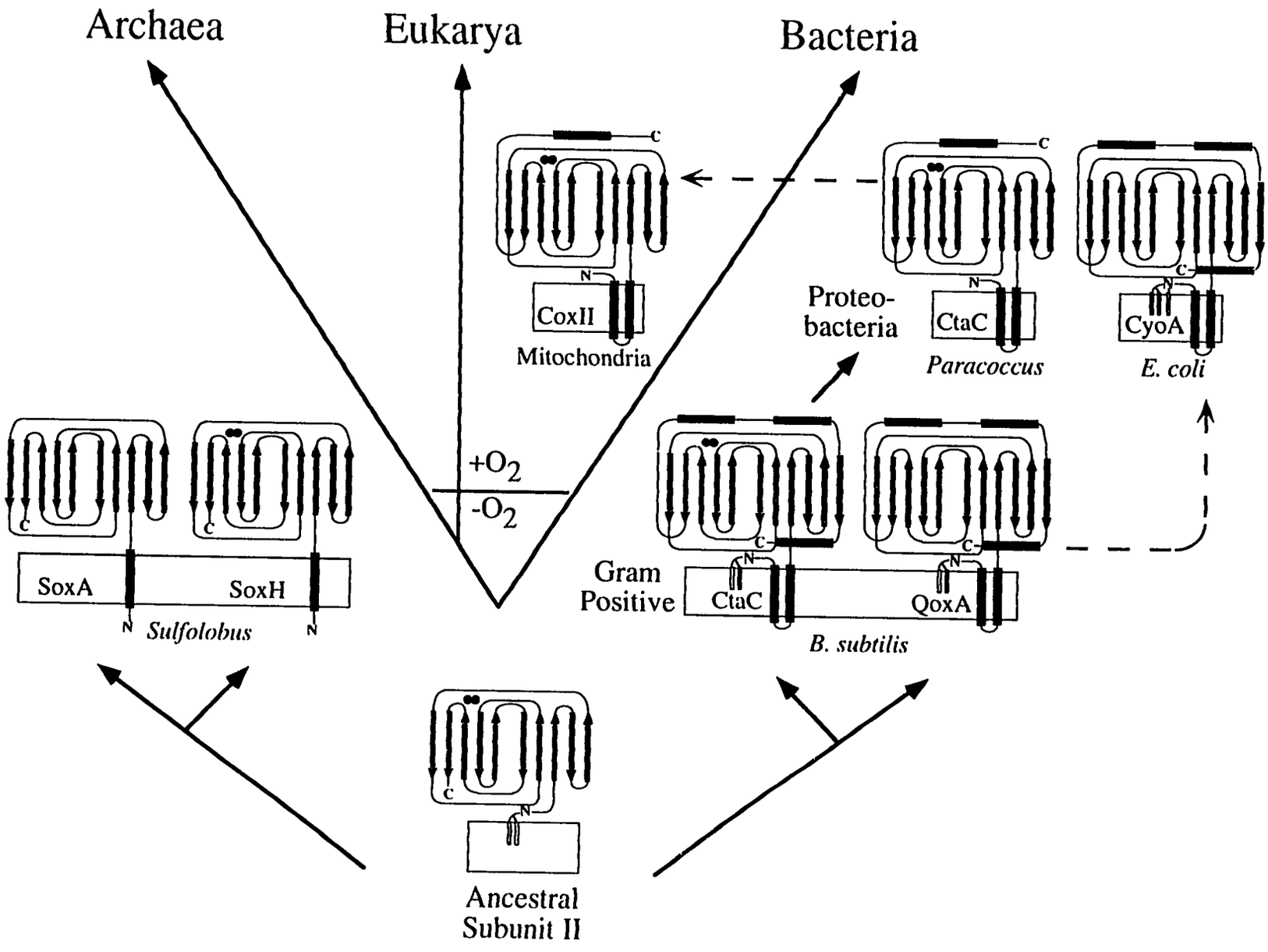


Figure 7-6 Growth of *E. coli* GO105 (*cyd, cyo*) carrying the wild type *cyo* operon on plasmid pRC07 (squares) or pC25A (circles), which carries a Cys25Ala mutation in the *cyoA* cistron. Duplicate colonies were cultured from plates of LB medium (1) supplemented with 0.3% D,L-lactate into corresponding liquid medium. Plating was staggered by 8 hours to ensure that colonies were the same size before culturing.

Figure 7-7:

Cartoon depicting a hypothetical scheme for the observed phylogenetic distribution of subunits II of cytochrome and quinol oxidases (21, 27), accounting for the presence of lipoprotein signal peptides (Table 7-1). The cupredoxin fold (gray arrows) carries an insertion of two ß-strands (black arrows), and can be further elaborated with C-terminal extensions that include α-helical regions (dark rectangles). Membrane spanning α-helical segments (black rectangles) at the N-terminus are anchored in membranes (open boxes). Diacylglycerylcysteine and an amide linked fatty acid (small boxes) may be present. The binuclear copper is indicated as tandem circles. Solid and dashed arrows indicate vertical and horizontal gene transfer events, respectively. The approximate time of O_2 accumulation in the atmosphere (28) is also indicated.


Archaea
Eukarya
Bacteria
CoxII
Mitochondria
Proteo-
bacteria
CtaC
Paracoccus
CyoA
E. coli
+O2
-O2
SoxA
SoxH
Sulfolobus
Gram
Positive
CtaC
QoxA
B. subtilis
Ancestral
Subunit II

References:

(1) Bishop, R.E. *et al.* (1995) *J. Biol. Chem.* 270, 23097-23103.

(2) Flower, D.R. *et al.* (1995) TiBS 20, 498-499

(3) Godovac-Zimmermann J. (1988) TiBS 13, 64-66

(4) Peitsch, M.C., and Boguski, M.S. (1990) *New Biol.* 2, 197-206

(5) Yang, C.Y. *et al.* (1994) *Biochemistry* 33, 12451-12455

(6) Ganfonina, M.D., Sanchez, D. and Bastiani, M.J. (1995) *Development* 121, 123-134

(7) Holden, H. *et al.* (1987) *EMBO J.* 6, 1564-1570

(8) Huber, R. *et al.* (1987) *J. Mol. Biol* 198, 499-513

(9) Timkovich, R., Cork, M.S., Gennis, R.B., and Johnson, P.Y. (1985) *J. Am. Chem. Soc.* **107:** 6069

(10) Provost, P.R. *et al.* (1991) *FEBS Lett.* 290,139-141

(11) Ding, Q., Kusano, S., Villarejo, M., and Ishihama, A. (1995) *Molec. Microbiol.* **16:** 649

(12) Seguin, D., Desforges, M. and Rassart, E. (1995) *Mol. Brain Res.* 30, 242-250

(13) Sanchez, D., Ganfonina, M.D. and Bastiani, M.J. (1995) *Development* 121, 135-147

(14) Rydén, L.G., and Hunt, L.T. (1993) *J. Mol. Evol.* **36:** 41

(15) Gotschlich, E.C., and Seiff, M.E. (1987) *FEMS Microbiol. Lett.* **43:** 253

(16) Kawula, T.H., Spinola, S.M., Klapper, D.G., and Cannon, J.G. (1987) *Molec. Microbiol.* **1:** 179

(17) Mattar, S., *et al.* (1994) *J. Biol. Chem.* **269:** 14939

(18) Saraste, M. (1990) *Quart. Rev. Biophys.* **23:** 331

(19) Wilmanns, M., *et al.* (1995) *Proc. Natl. Acad. Sci. USA* **92:** 11955

(20) Lindquist S., Weston-Hafer, K., Schmidt, H., Pul, C., Korfmann, G., Erickson, J., Sanders, C., Martin, H.H., and Normark, S. (1993) *Molec. Microbiol.* **9:** 703

(21) Castresana, J., Lübben, M., Saraste, M., and Higgins, D.G. (1994) *EMBO J.* **13:** 2516

(22) Ishizuka, M., *et al.* (1990) *J. Biochem.* **108:** 866

(23) Quirk, P.G., Hicks, D.B., and Krulwich, T.A. (1993) *J. Biol. Chem.* **268:** 678

(24) Fukaya, M., *et al.* (1993) *J. Bacteriol.* **175:** 4307

(25) Weyer, K.A., Schäfer, W., Lottspeich, F., and Michel, H. (1987) *Biochemistry* **26:** 2909

(26) Jixiang Ma , Andromachi Katsonouri, and Robert B. Gennis (1997) *Biochemistry* (submitted)

(27) Castresana, J., Lübben, M., and Saraste, M. (1995) *J. Mol. Biol.* **250:** 202

(28) Castresana, J., and Saraste, M. (1995) *Trends. Biochem. Sci.* **20:** 443

CHAPTER 8

The Entericidin Locus of E. coli: Implications for Programmed Bacterial Cell Death[1]

When I become death,

Death is the seed from which I grow.

William S. Burroughs

[1]A version of this chapter has been submitted for publication. Bishop, R.E., *et al.* (1997) *EMBO J.*

Introduction

Many components of the bacterial cell envelope are both essential and unique to the bacterium, and thereby provide important targets for antimicrobial chemotherapy and for the corresponding resistance mechanisms (Nikaido, 1994). The cell envelope of Gram-negative bacteria is a rigid structure that protects the bacterium from osmotic lysis and serves as a framework for the coordination of cell growth and division. The underlying structural component is the peptidoglycan or murein, which forms an exoskeleton that surrounds the cell and determines cell shape (Park, 1996). Extrinsic to the murein is the outer membrane, which is composed of an inner leaflet of phospholipid and an outer leaflet of lipopolysaccharide anchored by its lipid A substituent (Nikaido, 1996). The murein lipoprotein is covalently attached to the murein by its C-terminal lysine residue and anchored to the inner leaflet of the outer membrane by its lipid modified N-terminal cysteine residue (Wu, 1996). The outer membrane porins and several distinct lipoproteins provide additional non-covalent associations between the murein and the outer membrane (Rosenbush, 1974; Lecuc et al., 1992).

Despite the rigidity of the Gram-negative cell envelope, its individual components can be altered in response to specific environmental signals. For example, in response to osmotic stress, the expression of outer membrane porins can be modulated at the level of transcription by the EnvZ/OmpR two-component signal transduction pathway (Pratt and Silhavy, 1995; Harlocker et al., 1995). EnvZ is a membrane bound sensor-kinase that transfers a phosphoryl-group to OmpR, which controls transcription of porin genes. Starvation is also an important signal in the regulation of cell envelope components. The recent identification in E. coli of a starvation-inducible lipoprotein known as Slp (Alexander and St. John, 1994) was followed by the discovery of Blc, an outer membrane lipoprotein that is expressed at the onset of stationary phase under control of the *rpoS*-encoded stationary phase sigma factor σ^S (Bishop et al., 1995). Additionally, the OsmB

lipoprotein is governed by a σ^S-dependent promoter that is active only under high osmolarity conditions (Jung et al., 1990; Hengge-Aronis et al., 1991).

More than a dozen lipoproteins are known to be expressed from chromosomal genes in E. coli (Wu, 1996), but the genome probably encodes many more lipoproteins that have escaped detection because they are expressed at low levels or under specific environmental conditions. Lipoproteins represent a diverse group of unrelated proteins that share a common N-terminal type-II signal peptide. These can be identified by the presence of an invariant cysteine residue at the signal peptidase-II cleavage site (Wu, 1996). Signal peptides direct proteins to the cytoplasmic membrane, but only type-II signal peptides are modified by a thioether-linked diacylglyceryl substituent prior to cleavage of the signal peptide. The nascent α-amino group can then be modified by an amide linked fatty acid. Most lipoproteins are targeted to the inner leaflet of the outer membrane through a specific translocation mechanism (Matsuyama et al., 1995). However, some lipoproteins are retained on the outer leaflet of the inner membrane and these typically contain a negatively charged residue at the +1 or +2 position of the mature lipoprotein (Yamaguchi et al., 1988).

The deregulated synthesis of lipoproteins can have dire consequences for the bacterium. Two lytic-transglycosylases of E. coli are known to be lipoproteins (Ehlert et al., 1995; Höltje, 1995). These enzymes cleave the ß-1,4-glycosidic bond between N-acetylglucosamine and N-acetylmuramic acid in the murein, which is necessary for the insertion of new peptidoglycan material during cell growth and division (Höltje, 1995). However, lytic transglycosylases are potentially suicidal autolytic enzymes that can rupture the cell envelope if their activity is uncoupled from murein biosynthesis by ß-lactam antibiotics. Additionally, several plasmids of E. coli secrete toxins known as colicins that are normally released from the cell by the action of a colicin-release protein, which is a small bacteriolytic lipoprotein (van der Wal et al., 1995; Hsieh et al., 1997). Autolytic activity has been implicated in the programmed cell death of bacterial

populations (Ameisen, 1996). Since bacteria approach death stochastically in stationary phase, a moribund subpopulation could, in principle, support the remaining cells in the culture by releasing their cellular contents through bacteriolysis. Furthermore, programmed bacterial cell death mechanisms have been identified in the context of plasmid addiction modules, which typically comprise an antidote/toxin gene pair (Yarmolinsky, 1995). We now report a novel chromosomal bacteriolytic module of enterobacteria called the entericidin locus. The entericidins are small lipoproteins encoded by tandem genes, which function as an antidote/toxin pair that is under positive control from σ^S and negative control from EnvZ/OmpR.

Results

Analysis of the entericidin locus

The entericidin locus (*ecn*) occupies a 500 bp segment downstream of the gene that encodes elongation factor-P (Aoki et al., 1991) and resulted from an apparent gene duplication event, which predated the evolutionary divergence of E. coli and Citrobacter freundii (Figure 1A,B). In the latter species, the tandem genes are displaced from the downstream *sugE* gene by a divergently oriented open reading frame (*ecnR*), which specifies a member of the response regulator family of transcription factors. The closest relative of EcnR is RedZ (Figure 1C), which is reported to control antibiotic production in Streptomyces coelicolor (GenBank Y07902). The amino acid sequence of C. freundii SugE has been reported previously (Paulsen et al., 1996) where it was identified as a member of the SMR family of small multidrug resistance efflux pumps.

The deduced amino acid sequences of the entericidins exhibit type-II signal peptides characteristic of bacterial lipoproteins (Figure 1B) and are similar in size to the colicin release proteins. These are known to induce a quasi-lysis phenomenon characterized by a gradual decrease in cell density during colicin production (van der Wal et al., 1995). However, no significant amino acid sequence similarity is shared between the colicin release proteins and the tandem lipoproteins, which we have termed entericidin A (EcnA) and entericidin B (EcnB). An outer membrane localization of the entericidins is predicted from the absence of a negatively charged residue at the +1 or +2 position, which is a cytoplasmic membrane localization signal for E. coli lipoproteins (Yamaguchi et al., 1988). An E. coli mutant (MI1443) deleted in a chromosomal segment that spans the entericidin locus is viable (Bishop and Weiner, 1993b), indicating that the entericidin locus is non-essential. Furthermore, the entericidin locus appears to be unique among the Enterobacteriaceae as it is absent from the currently completed genomes of other bacterial species.

The *ecnA* gene does not exhibit any obvious transcriptional regulatory signals and lies only 20 nucleotides downstream of the putative Rho-independent terminator for the upstream operon which encodes elongation factor-P (Figure 1A). When the colicin release proteins are similarly encoded behind Rho-independent terminators, they can be expressed by translational coupling with the preceding genes that encode the colicin and its immunity protein (Hsieh et al., 1997). A potential σ^S promoter element preceding the *ecnB* gene is identical to the promoter of *osmB* (Figure 2A). The σ^S consensus sequence comprises a heptad element plus an upstream region of intrinsic DNA curvature (Espinosa-Urgel et al., 1996). Intrinsic DNA curvature was identified in the corresponding region by the E. coli genome project (Burland et al., 1995), which overlooked the small entericidin open reading frames. A sequence resembling a Rho-independent terminator lies 230 nucleotides downstream of the *ecnB* promoter and a good ribosome binding site lies upstream of the putative initiating methionine codon (Figure 1A). Additionally, a tandem repeat characteristic of an OmpR binding site (Pratt and Silhavy, 1995; Harlocker et al., 1995) lies immediately downstream of the *ecnB* promoter element, but only in E. coli (Figures 1A, 2A).

Transcriptional regulation of the entericidin locus

Northern blot and primer extension studies demonstrate that the E. coli *ecnB* transcript is synthesized in vivo and begins from the predicted promoter element (Figure 2B, C). Total cellular RNA was isolated from E. coli cultures grown to exponential phase (180 min of growth) or early stationary phase (360 min of growth) in Luria-Bertani medium supplemented with and without 300mM NaCl. The E. coli strains included wild-type, an *rpoS359*::*Tn*10 derivative, an *envZ60*::*Tn*10 derivative, and an *ompRenvZ*::*Tn*10 derivative. When northern blots of stationary phase RNA were probed with a ^{32}P-end-labeled primer derived from the sequence of *ecnB*, a signal corresponding to a transcript of approximately 250 nucleotides was detected optimally in the wild-type strain under high salt growth conditions, whereas no signal was identified in the *rpoS359* mutant. The

same signal was fully derepressed in the *envZ60* and *ompRenvZ* mutants regardless of osmolarity (Figure 2B). 23S and 16S ribosomal RNA from the stationary phase samples were visualized by cross hybridization with a labeled digest of DNA size standards derived from pBR322 and showed that the same relative amounts of RNA were present in each lane (Figure 2B). Although the *ecnB* transcript was not detected with exponential phase RNA from the wild-type strain or the *rpoS359* mutant, a weak signal was detected from the *envZ60* and *ompRenvZ* mutants (not shown). The accumulation of the *ecnB* transcript in the *envZ60* and *ompRenvZ* mutants under low salt conditions, and in exponential phase, resembles the derepression of other *rpoS*-dependent promoters attributed to various ancillary regulatory factors (Hengge-Aronis, 1996a,b). When the *ecnB* probe was used in primer extension studies with RNA from the NaCl-supplemented stationary phase cultures, a doublet of bands, which paralleled the northern blot expression levels, confirmed that the transcriptional start site corresponds with that predicted from the *osmB* promoter (Figure 2C).

The entericidin locus governs stationary phase autolysis

Wild-type E. coli cultures carrying low copy number plasmids bearing the *frd-efp* gene intervals from E. coli or C. freundii (Figure 3A) were found to undergo a gradual clearing in stationary phase when cultured aerobically in Luria-Bertani medium. Examination by microscopy indicated that the cells in the cultures had lysed. We traced the bacteriolysis phenotype to the entericidin locus and subcloned the entericidin genes, individually and in tandem, into an IPTG-inducible *tac* promoter expression vector (Figure 3B). In the absence of IPTG, the tandem C. freundii genes in pCfEcnAB exhibited a bacteriolysis phenotype that was accentuated compared to that seen previously with pAmpRC (see below). Under the same conditions, bacteriolysis was no longer observable with the endogenous E. coli entericidin locus in pEcEcnAB owing to repression from EnvZ/OmpR. The gradual bacteriolysis observed in both pAmpAC and pAmpRC appeared to depend on the *cat*-promoter of the vector (Figure 3A). Although it

remains to be determined whether C. freundii EcnR serves to repress the entericidin locus in this organism, the difference in regulation provided a useful assay to monitor the contribution of each entericidin to stationary phase autolysis. The C. freundii plasmids (Figure 3B) were transformed into wild-type E. coli and cells were cultured in Luria-Bertani medium supplemented with and without 300mM NaCl, and with or without 1mM IPTG to induce the *tac* promoter. A representative experiment, summarizing routinely obtainable findings, is shown in Figure 4.

Under high salt conditions in the absence of IPTG, bacteriolysis only occurs at the onset of stationary phase when both entericidin genes are intact and *ecnB* is under control of its endogenous σ^S promoter. As expected, bacteriolysis was largely eliminated when the cells harboring both entericidin genes were grown in the absence of salt. Under high salt conditions in the presence of IPTG (Figure 4), bacteriolysis was most pronounced when the lone *ecnB* gene was activated. This was also true in the absence of salt supplement, indicating that *ecnB* expression ultimately determines bacteriolysis. However, because IPTG was added at time zero, and bacteriolysis ensued only in late exponential phase, the result further indicates that exponentially growing bacteria are immune somehow to the lytic action of EcnB. It appears that the *ecnB* promoter is regulated to coordinate EcnB production with the exact moment the cell is susceptible to undergo bacteriolysis. Although expression of *ecnA* alone had no effect on cell growth, the expression of *ecnA* in *cis* to *ecnB* served to suppress bacteriolysis in the presence of salt (Figure 4); the remaining gradual bacteriolysis is comparable to the earlier observations with pAmpAC and pAmpRC. We find that expression of *ecnA* in *trans* to *ecnB* also serves to suppress bacteriolysis under the same conditions (not shown). This demonstrates that bacteriolysis is specific to the product of *ecnB* and that a diffusible antidote is encoded by *ecnA*, which can be expressed from a promoter located upstream of the preceding Rho-independent terminator. By analogy with the colicin operons (van

der Wal et al., 1995; Hsieh et al., 1997), the elongation factor-P promoter may serve this function in vivo.

The entericidins are membrane lipoproteins

We have confirmed the lipoprotein nature of the *ecnA* and *ecnB* gene products in two ways. First, we expressed the genes in the presence of ^{3}H-palmitate and analyzed cellular protein with a Tricine-SDS-PAGE system designed for the resolution of small peptides (Schägger and von Jagow, 1987; Sankaran et al., 1995). The cloned *ecnA* gene from E. coli was analyzed, because this construct (pEcEcnA; Figure 3B) lacks the upstream Rho-independent terminator that attenuates EcnA production. The results of fluorography demonstrate that expression of EcnB is at least an order of magnitude greater than that of EcnA (Figure 5A), consistent with the poorer ribosome binding site preceding the *ecnA* gene (Figure 1A). The production of EcnB compares with the abundant murein lipoprotein (Mlp). When C. freundii EcnB was expressed in the presence of globomycin, an inhibitor of the lipoprotein signal peptidase (Hussain et al., 1980), a larger species, corresponding to the diacylglyceryl-prolipoprotein derivative of EcnB, was detected (Figure 5B). This experiment was unsuccessful when EcnA was analyzed, perhaps owing to its weaker expression or instability. Second, we detected the entericidins in membranes by silver staining (Figure 5C). As predicted from the results of fluorography, at least an order of magnitude greater expression was obtained with EcnB as compared with EcnA. However, both entericidins were detected in membranes in similar amounts when they were co-expressed in tandem (Figure 5C).

Conformational analysis of soluble entericidin derivatives

Small amphipathic α-helical peptides are known to modulate membrane bilayer properties (Epand et al., 1995; Tytler et al., 1993). Intriguingly, the entericidins also exhibit amphipathic features when they are modeled in an α-helical conformation starting from the putative lipoprotein cleavage site (Figure 6A). In order to determine whether the entericidins have the potential to adopt α-helical structures, we synthesized

entericidins with the N^{α}-acyl-S-diacylglycerylcysteine residues replaced by N^{α}-acetylserine. The peptides were highly soluble in benign buffer (50 mM potassium phosphate pH 7 containing 50 mM KCl at 25C) and exhibited predominantly random coil properties as determined by far ultraviolet circular dichroism (CD) analysis (Figure 6B). However, in the presence of a membrane mimetic solvent (50% trifluoroethanol, TFE), α-helical structure is indicated by the shift in the minimum ellipticity peak from 197 to 208 nm ($\pi\pi^*$-transition) and the negative ellipticity at 222nm ($n\pi^*$-transition). The helicity calculated from the algorithm of Greenfield and Fasman (1969) is low for the entericidins in benign buffer (19% for EcnA and 14% for EcnB), but the values increase in the presence of TFE (59% for EcnA and 38% for EcnB). The helical content calculated from secondary structure prediction algorithms (Wishart et al., 1994) more closely resemble the values determined in the presence of TFE (65% for EcnA and 30% for EcnB). A strong correlation between TFE-induced and predicted helicity in studies of a synthetic actin peptide was previously noted (Sönnichsen et al., 1992).

Discussion

The proliferation of unicellular organisms is controlled by the rates of biochemical reactions that comprise the cell cycle and are ultimately poised by the exogenous nutrient supply. Some unicellular organisms respond to nutrient deprivation by activating a genetic program of sporulation, which allows the cell to remain dormant during protracted periods of starvation. However, unicellular organisms can also adapt to starvation by adjusting the size of their populations by means of programmed cell death (Ameisen, 1996). If the activation of programmed cell death in unicellular populations were an all-or-none response, the effects would be irrevocable, but the elimination of a moribund subpopulation by a stochastic genetic switch could prove beneficial by providing nutrients to the remaining healthy cells in the population. Additionally, the mechanisms of programmed cell death in unicellular organisms may have provided evolutionary antecedents for programmed cell death in multicellular organisms.

In this study, we have discovered and characterized a novel chromosomal programmed cell death module of enterobacteria, which we have named the entericidin locus. The entericidin locus resulted from a gene duplication event, which predated the evolutionary divergence of E. coli and C. freundii and is presumably conserved in other enterobacteria. The entericidin locus was revealed by a stationary phase bacteriolysis phenotype associated with the cloned C. freundii locus when it was carried in E. coli. This resulted from a derepression owing to an absence of negative control from the EnvZ/OmpR two-component signal transduction pathway, which regulates the endogenous entericidin locus of E. coli. The E. coli *ecnB* gene appears to provide the first example of classical repression mediated by an *ompR* operator located downstream of its cognate promoter, but the possibility remains that repression may occur by OmpR binding upstream of the promoter at sites not readily apparent (Pratt and Silhavy, 1995b; Harlocker et al., 1995). The C. freundii entericidin locus is linked to an OmpR homolog called EcnR that may provide repression in this organism. EcnR is only three open

reading frames removed from AmpR (Bishop and Weiner, 1993a; Jacobs et al., 1997), which is a C. freundii transcriptional regulator that has been similarly deleted from E. coli (Figure 3A).

The *ecnB* promoters in both E. coli and C. freundii belong to a subclass of σ^S promoters that depend on the cytoplasmic accumulation of potassium glutamate in response to high extracellular salt concentrations (Ding et al., 1995). The *ecnB* promoter of E. coli is unique in that its activation in response to osmolarity is checked by osmotic repression due to the EnvZ/OmpR signal transduction pathway. This seemingly paradoxical regulation may provide a stochastic genetic switch that responds to a threshold signal to govern bacteriolysis of a moribund subpopulation. Although the *ecnB* transcript of E. coli was detected in vivo, the absence of observable bacteriolysis could be due to a rapid replacement of any lysed cells. Compensatory changes in the distribution of outer membrane porins may account for the viability of *ompRenvZ* mutants.

Bacteriolysis globally affected wild-type populations when the entericidin locus was derepressed, but E. coli cultures do not normally do this at the onset of stationary phase. However, as a result of prolonged starvation, E. coli populations can be replaced by mutants that have a growth advantage in stationary phase known as the "GASP" phenotype (Zambrano et al., 1993; Zambrano and Kolter, 1996) and some of the responsible mutations have been traced to *rpoS*. GASP has been interpreted to result from an ability of the mutants to "cannibalize" the wild-type cells. In principle, GASP mutants could passively replace the population if they have simply lost the ability to undergo a normal bacteriolysis during starvation. While the entericidin locus can provide *rpoS*-dependent bacteriolysis, it remains to be determined whether it affects the GASP phenotype. This may provide a considerable challenge in light of the fact that the entericidin locus is not the only programmed cell death module on the E. coli chromosome.

Programmed cell death in bacterial populations was initially identified in the context of plasmid maintenance systems known as "addiction modules" (Yarmolinsky, 1995; Jensen and Gerdes, 1995). These encode antidote/toxin gene pairs that are regulated to activate the toxin on plasmid loss, thereby maintaining the plasmid in the population. Intriguingly, several homologs of plasmid addiction modules have been identified in a chromosomal context and one of these, E. coli *mazEF*, is the downstream component of the *relA* operon (Aizenman et al., 1996). RelA is a ribosomal subunit that senses starvation and responds to uncharged aminoacyl-tRNAs by synthesizing guanosine-3',5'-bispyrophosphate (ppGpp), which is an intracellular starvation signal of the stringent response. High levels of ppGpp downregulate the transcription of *mazEF*, which unleashes the toxin from its labile antidote and kills the cells (Aizenman et al., 1996). The killing mechanism is not intrinsically bacteriolytic and the homologous plasmid-borne toxins target the machinery of DNA replication (Jensen and Gerdes, 1995).

The entericidin locus and the *mazEF* addiction module exhibit some intriguing similarities and differences. Both loci are encoded immediately downstream of a component of the translational apparatus and comprise an antidote/toxin gene pair that programs cell death under starvation conditions (Figure 7). However, the entericidin locus has no known derivatives on plasmids, nor is its killing mechanism likely to exert a post-segregational killing effect in a plasmid context because cell death results from the active synthesis of a toxin rather than the passive degradation of a labile antidote. The antidote function of *ecnA* may simply provide a guard against inappropriate transcription of *ecnB*.

The entericidins may operate like the colicin release proteins, which have been implicated in the control of an outer membrane phospholipase (Pugsley and Schwartz, 1984). However, since non-pathogenic E. coli strains are not known to secrete proteins (Pugsley, 1993), it seems unlikely that the entericidins are part of a secretory pathway.

The presumed amphipathic α-helical nature of the membrane bound entericidins is revealing, because similar motifs are known to differentially modulate membrane bilayer properties (Epand et al., 1995). It has been suggested that the cross-sectional molecular shapes of apolipoprotein and lytic peptide analogs determine reciprocal control of membrane stability (Tytler et al., 1993). However, the reciprocal effects of the entericidins on the Gram-negative cell envelope may involve distinct physical principles owing to the combination of the amphipathic α-helix and lipoprotein modification. In this regard, the distribution of residues in EcnA, which is an antidote to the lytic EcnB, more closely resemble lytic peptides than apolipoproteins (Figure 6A; Epand et al., 1995).

Like other soluble membrane active peptides (Zagorski et al., 1991), soluble synthetic entericidins adopt random coil configurations in water, but assume α-helical structure in the presence of 50% trifluoroethanol, which provides a dielectric constant that approximates conditions at the interface between the lipid bilayer and the aqueous phase. Secondary structure induced by lipidation of peptides (Macquaire et al., 1992) suggests that the entericidins bearing N^{α}-acyl-S-diacylglycerylcysteine in vivo may contain greater secondary structural propensity than that determined by circular dichroism for the N^{α}-acetylserine derivatives. The greater degree of helicity exhibited by EcnA, which contains fewer glycine residues than are found in EcnB, may provide a structural basis for the differential effects of the entericidins. In this regard, a flexible hinge is predicted at the tandem glycine residues +14 and +15 of EcnB (Figure 1B).

The entericidin genes resemble the products of an antimicrobial peptide-encoding gene in which the secretory signal peptide was converted to a lipoprotein signal peptide that retains the product in the cell envelope. This supports the recent proposal that programmed cell death genes may have originated in bacteria from a pool of antibiotic genes (Ameisen, 1996). The duplication of the ancestral entericidin gene is also consistent with the proposal that the ancestor of the apolipoprotein multigene family

evolved from a primordial minigene which was duplicated to give rise to the third and fourth exons that encode the amphipathic α-helical segments (Boguski et al., 1986). The entericidin locus of E. coli is only one transcriptional unit removed from the gene that encodes Blc (Figure 3A), the bacterial ancestor of the lipocalin known as apolipoprotein D, which first demonstrated homology between an outer membrane lipoprotein of enteric bacteria and an apolipoprotein component of plasma lipoproteins (Bishop et al., 1995). Gram-negative outer membranes and plasma lipoproteins represent two rare examples in biology of asymmetric lipid bilayers that comprise a single leaflet of phospholipid.

Intriguingly, the entericidins also share design features with the colicin release proteins of colicin plasmids (van der Wal et al., 1995; Hsieh et al., 1997). Although colicins and their immunity determinants have long been regarded as components of cellular defense, they can also be regarded as products of an addiction module engaged in colicin plasmid maintenance. The recent finding that a structural and functional similarity exists between the pore-forming colicins and Bcl-x_L, which regulates programmed cell death in multicellular eukaryotes (Muchmore et al., 1996), serves to underscore the likelihood that the machinery of programmed cell death is conserved between diverse organisms. As in multicellular organisms, the mechanisms of programmed cell death in bacteria depend on the balanced expression of genes that promote either survival or destruction in response to appropriate cellular signals. Further consideration of the role that the entericidin locus plays in regulating bacteriolysis of stationary phase bacteria, which are intrinsically resistant to many antibacterial agents, may be of future benefit in the design of novel antibiotics.

Experimental Procedures

Bacterial strains and plasmids

E. coli MC4100 (wild type) and RH90 (MC4100 *rpoS359::Tn*10) have been described previously (Bishop et al., 1995), as were LP330 (MC4100 *envZ60::Tn*10) and LP64 (MC4100 *ompRenvZ::Tn*10) (Pratt and Silhavy, 1994), and MC1061 (Casadaban and Cohen, 1980). Chromosomal DNA was derived from plasmids pAmpAC (E. coli; Bishop et al., 1995) and pNU302 (C. freundii; Lindberg et al., 1985) from which the 7.6 kb EcoRI fragment was moved into pACYC184 to create pAmpRC. 12.5 mL cultures were grown in 125mL Klett flasks and monitored as described previously (Bishop et al., 1995).

The pMS119 expression vectors carry an isopropyl-ß-D-thiogalactopyranoside (IPTG) inducible *tac*-promoter preceding a multiple cloning site, in two possible orientations (pMS119EH and pMS119HE), followed by the *rrnB* transcriptional terminator (Strack et al., 1992). A 1.3-kb EcoRI/SphI fragment from pAmpRC was cloned into pMS119EH to create pCfEcnAB; the 0.7-kb EcoRI/BglII fragment was cloned into the EcoRI/BamHI digested pMS119EH to create pCfEcnA; the 0.5-kb BglII/SphI fragment was cloned into the BamHI/SphI digested pMS119EH to create pCfEcnB. The 0.5-kb PstI/PvuII fragment (Figure 1A,3) from pAmpAC was cloned into the PstI/SmaI digested pMS119HE to create pEcEcnAB. The 0.2-kb HindIII/AseI fragment from pEcEcnAB was cloned into the HindIII/NdeI digested pET-22b (Novagen) and the HindIII/XbaI fragment moved back into pMS119HE to create pEcEcnA. The 0.2-kb AseI/EcoRI fragment from pEcEcnAB was cloned into the NdeI/EcoRI digested pET-22b and the XbaI/EcoRI fragment moved back into pMS119HE to create pEcEcnB.

Sequence analysis

Plasmid purifications, restriction enzyme digestions, ligations, transformations, and DNA electrophoresis were performed as described (Sambrook et al., 1989). DNA synthesis and sequencing was performed as described previously (Bishop et al., 1995) in

the Department of Biochemistry Core Facility (University of Alberta) using an Applied Biosystems 392 DNA/RNA synthesizer and 373A DNA sequencer (Perkin Elmer, Foster City CA, U.S.A.). A portion of the entericidin locus was unknowingly characterized in a previous investigation of *sugE* (Greener et al., 1993), which identified an amino terminal extension in E. coli CS520 known as *sugEL*. However, this sequence was not congruent with that determined by the E. coli MG1655 genome project (Burland et al., 1995). Our clone in plasmid pAmpAC (Bishop et al., 1995) was also derived from E. coli CS520, and we found that the sequencing error is due to the absence of T 385 (Figure 1A), which disrupts *ecnB*, and AAT 491-3 (Figure 1A), which creates the *sugEL* extension. The DNA sequences reported here will be updated in GenBank files U21726 (E. coli) and U21727 (C. freundii). Translated amino acid sequences were compared with GenBank using the BLAST algorithm (Altschul et al., 1990). The HELICALWHEEL (Schiffer and Edmundson, 1967) and GAP algorithms were used from the Genetics Computer Group (University of Wisconsin).

Northern analysis and transcriptional mapping

RNA was prepared as described (Bishop et al., 1995) and digested with RNase-free DNase I (Boehringer Mannheim; this was also the source of low molecular mass DNA size standards from pBR322). Northern blotting was performed as described (Williams and Mason, 1985). 40 μg of glyoxal denatured RNA was separated on a 1.25% agarose gel for capillary blotting and probing. Primer extension was performed as described (Bishop et al., 1995) where nucleotide sequence was determined with the Sequenase kit (U.S. Biochemical). The EcnB primer was synthesized and labeled with [γ-^{32}P]ATP as described (Bishop et al., 1995) and hybridizes between nucleotides 400 and 419 (Figure 1A): 5'-ACGCGTGGTGTTGCAGGCAG-3'.

Lipoprotein characterization

Palmitoylation studies were performed as described previously (Bishop et al., 1995), except that the Tricine-SDS-PAGE system was utilized (Schägger and von Jagow,

1987; Sankaran et al., 1995). The silver staining protocol was developed by R.J. Turner and S. Brown. Gels were fixed for 1 hr in 25% isopropanol/7% acetic acid then oxidized for 5 min in 0.1% $K_2Cr_2O_7$/0.2% HNO_3 followed by four 1 min washing steps in distilled water. The gel was then soaked in 0.1% $AgNO_3$ for 5 min on a light box, followed by 30 min on the bench. A brief wash in distilled water was followed by development in 3% Na_2CO_3/0.05% formaldehyde as required. Development was terminated upon addition of 25 mL 2.3M citric acid.

Peptide analysis

EcnA (23 residues) and EcnB (27 residues) in which the amino terminal N^{α}-acyl-S-diacylglycerylcysteine residues are replaced by N^{α}-acetylserine were synthesized by solid phase procedures using a N^{α}-Boc amino acid PAM resin, which provides a C-terminal free acid after cleavage from the resin, on an Applied Biosystems model 430A peptide synthesizer (Foster City, CA) using N^{α}-t-butyloxycarbonyl chemistry as reported (Lavigne et al., 1995). The same procedures (Lavigne et al., 1995) were used for N^{α}-acetylation, purification by reversed-phase liquid chromatography, amino acid analysis, and electrospray mass spectrometry.

CD spectra were measured on a Jasco J-500C spectropolarimeter (Jasco, Easton, MD) equipped with a Jasco DP-500N data processor as described (Lavigne et al., 1995), where ellipticity is reported as mean residue ellipticity ([θ]) in deg cm^2 $dmol^{-1}$. The estimation of peptide conformation from CD spectra was determined by the method of Greenfield and Fasman (1969) using the software package CDPENCE written by R. Boyko and D. Wishart, University of Alberta. Secondary structure predictions were made with the software package SEQSEE (Wishart et al., 1994).

Figure 8-1. DNA Sequence of the Entericidin Locus from E. coli CS520 and the Deduced Amino Acid Sequences of the Entericidin Open Reading Frames.

(A) The final 18 codons of the *ef-P* gene are followed by a Rho-independent terminator (underlined) flanking a PstI restriction site. The 41 codons of *ecnA* are preceded by a ribosome binding site (rbs) and then followed by the RpoS promoter element of *ecnB*, an AseI restriction site, the *ecnB* transcriptional start site (+1), and a tandem repeat (>>>) characteristic of a putative OmpR binding site. The 48 codons of *ecnB* are preceded by a ribosome binding site and then followed by a putative Rho-independent terminator (underlined) flanking a PvuII restriction site. The cysteine residues at the putative type-II signal peptidase cleavage sites are outlined. Nucleotides in bold type were erroneously absent from an earlier analysis of *sugE* (Greener et al., 1993).

(B) The entericidins from Citrobacter freundii (CfEcnA and CfEcnB) are aligned with their homologs from E. coli (EcEcnA and EcEcnB). The positive charges at the start of the type-II signal peptides are marked (++) as are the cysteine residues (outlined) at the putative lipoprotein processing sites (+1). Identical residues are emphasized in bold type.

(C) Alignment of EcnR from C. freundii with RedZ from Streptomyces coelicolor. The alignment was produced with the GAP algorithm (Genetics Computer Group, Wisconsin). Percent Similarity: 44.444. Percent Identity: 27.513.

A

```
E  V  I  K  V  D  T  R  S  G  E  Y  V  S  R  V  K  *
GAAGTCATCAAAGTGGATACCCGCTCTGGTGAATACGTCTCTCGCGTGAAGTAATGCGGT 60
                                                 EcnA>
                PstI                        rbs      M  M  K  R  L   5
TGTGGTGCGGCCTGCAGGCTGCACCATCACTTATTCAGGTCAGAGATGATGAAACGCCTT 120

I  V  L  V  L  L  A  S  T  L  L  T  G  C  N  T  A  R  G  F   25
ATCGTTCTTGTTTTGCTTGCCAGCACGCTGCTCACGGGCTGTAACACCGCTCGCGGTTTC 180

G  E  D  I  K  H  L  G  N  S  I  S  R  A  A  S  *            41
GGCGAAGACATCAAACATCTCGGCAACTCCATCTCTCGCGCTGCCAGCTAATTTTTCTTC 240

                                                    RpoS AseI +1
TCTTCCGAAAAATCATCAGATTCCCATCATTTTTGGCGATGTTGTCTATTATTAATTTGC 300
                OmpR                                EcnB>
          >>>>>    >>>>>        rbs        M  V  K  K  T  I  A  7
TATAGGCAAACATAAATAACATTACCTAAAAGGAAGACGTTATGGTGAAGAAGACAATTG 360

  A  I  F  S  V  L  V  L  S  T  V  L  T  A  C  N  T  T  R  G  27
CAGCGATCTTTTCTGTTCTGGTGCTTTCAACAGTATTAACTGCCTGCAACACCACGCGTG 420

  V  G  E  D  I  S  D  G  G  N  A  I  S  G  A  A  T  K  A  Q  47
GCGTTGGTGAAGACATTTCTGATGGCGGTAACGCGATTTCTGGCGCAGCAACGAAAGCGC 480

  Q  *                        PvuII                            48
AGCAATAAGCAATAACGGTACGACAGCTGTGTCGTGCCGTTTGTTTTTTCTGCGATAGTC 540
```

B

```
                ++                  +1
CfEcnB   MVKKTIAAIFSVLVLSSVLTA-CNTTRGVGQDISEGGSAISGAGTKAQQ
EcEcnB   MVKKTIAAIFSVLVLSTVLTA-CNTTRGVGEDISDGGNAISGAATKAQQ
CfEcnA   MMKRLLGLVMLLLFTCTLLTG-CNTARGFGEDIKHLGNSISHAAS----
EcEcnA   MMKRLIVLVLL---ASTLLTG-CNTARGFGEDIKHLGNSISRAAS----
```

C

```
RedZ    1 MTTRVLVCCDRVILGEGIRALL...ERHDMKVQVETTQRGSLATAAETGP 47
            ::||.    :  .|| |||   :| . .. |.....  ||..   .
EcnR    1 .MLKILVIDRCHFTRTGIEALLNHSGRFSSSFLVSGINNLLLAKEHILQW 49

RedZ   48 DILVGVAPLFTMDS.......IDKLTELARLGKTLLLTKPENTHRAFEAL 90
          .  :.:|.| .: |       |:.:   . : . :|| .::. | ::.:
EcnR   50 KPHLVIADLNSFISETHSSPPINPFFMSCGVIPLILLQSADRQHAPIAPS 99

RedZ   91 RVGVRAVLSAETSVEELVHVIRTITEVNAIIAPEEAQEALTRYWRSKAPK 140
          . ..:.||. .|.::.| |.|..  :|.: :               . ||
EcnR  100 QSVAHSVLTKHTTLNTLSHTIQEALQVRPAL..............EIPK 134

RedZ  141 NLRPELTPRETEVLLLLTQGKTNTEMAATLSVSPTTVRSHVHRILRKLGA 190
          |  | |||.|..|| :: :| .|..:||.||: ..|| .. :.| .||
EcnR  135 NATPLLTPQEEKVLSMWMDGVSNNAIAAALSIHGKTVYTYKRNIRMKLHL 184

RedZ  191 ATRAQAVAIAYESGLLGICPGYGTPAR 217
          :.|         |.:|:: || |.
EcnR  185 GNRF........SPFLSL.PGKGD... 199
```

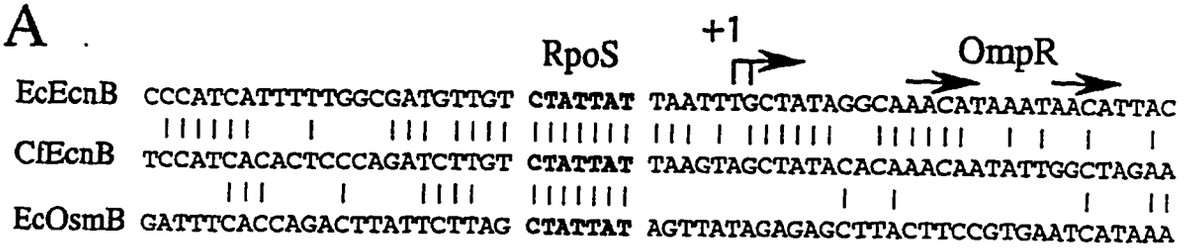
Figure 8-2. Transcriptional Mapping and Regulation of the *ecnB* Promoter.

(A) The RpoS promoter elements from the *ecnB* genes of E. coli (EcEcnB) and C. freundii (CfEcnB) are aligned with that of the E. coli *osmB* gene (EcOsmB) (Jung et al., 1990; Hengge-Aronis, 1996a,b). The transcriptional start site for EcEcnB is marked (+1) as is the tandem repeat characteristic of an OmpR binding site (Pratt and Silhavy, 1995; Harlocker et al., 1995), which is only present in EcEcnB.

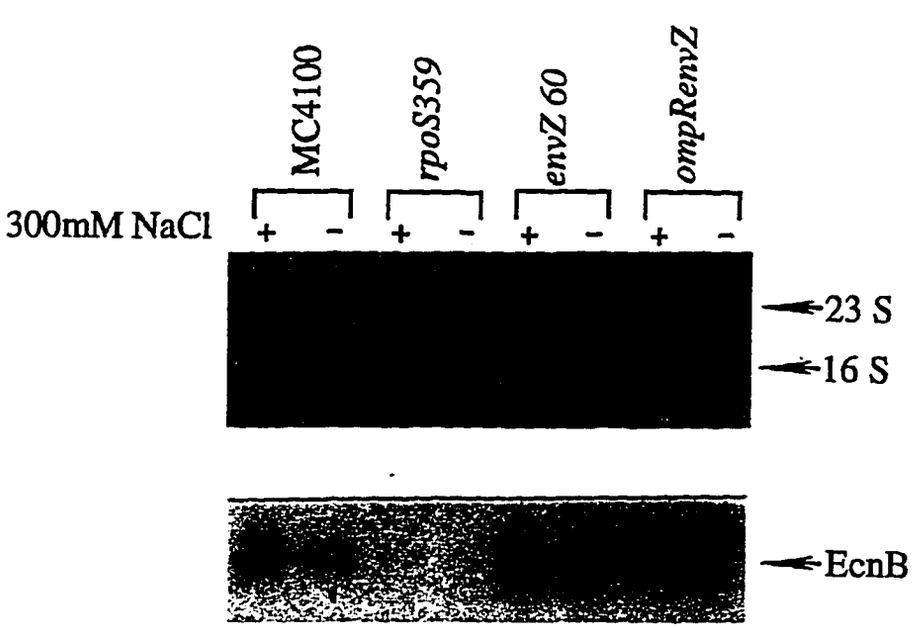
(B) Northern analysis of total cellular RNA from E. coli MC4100 and its *rpoS359*, *envZ60*, and *ompRenvZ* mutant derivatives; these were grown to stationary phase in Luria-Bertani medium supplemented with or without 300mM NaCl. The signal detected with the ^{32}P-labeled *ecnB* oligonucleotide is shown in the bottom panel. The blot was then stripped and re-probed with ^{32}P-labeled DNA markers from pBR322, which hybridized with the 23S and 16S ribosomal RNAs (top panel).

(C) Primer extension analysis was performed with the ^{32}P-labeled *ecnB* oligonucleotide hybridized to the RNA samples from E. coli MC4100 and its *rpoS359*, *envZ60*, and *ompRenvZ* mutant derivatives grown to stationary phase in the presence of 300mM NaCl. The deduced sequence of the *ecnB* promoter derived from the sequencing lanes labeled C, T, A, and G is shown vertically on the left. The position of the cDNA corresponding to the *ecnB* transcriptional start site (+1) is marked by a pair of arrowheads.

A


RpoS
+1
OmpR
EcEcnB CCCATCATTTTTGGCGATGTTGT CTATTAT TAATTTGCTATAGGCAAACATAAATAACATTAC
CfEcnB TCCATCACACTCCCAGATCTTGT CTATTAT TAAGTAGCTATACACAAACAATATTGGCTAGAA
EcOsmB GATTTCACCAGACTTATTCTTAG CTATTAT AGTTATAGAGAGCTTACTTCCGTGAATCATAAA

B


MC4100
rpoS359
envZ 60
ompRenvZ
300mM NaCl
+ - + - + - + -
23 S
16 S
EcnB

C

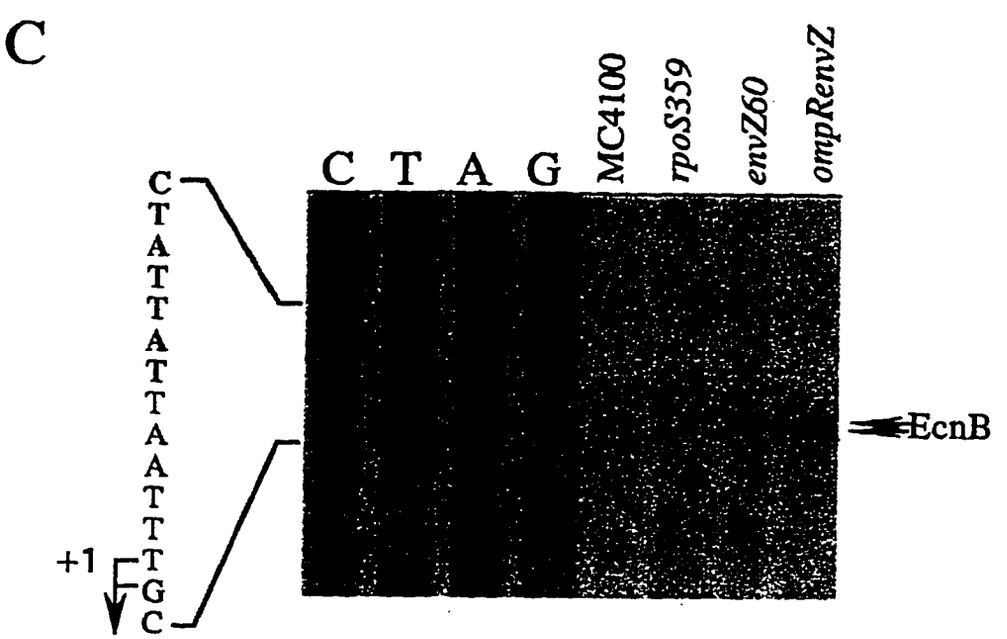

C T A G
MC4100
rpoS359
envZ60
ompRenvZ
C T A T T A T T A A T T T G C
+1
EcnB

Figure 8-3. Cloning of the Entericidin Genes.

(A) The *frd-ef-P* gene interval of E. coli is carried by a 6.1 kb EcoRI fragment in pAmpAC (Bishop et al., 1995). The homologous 7.6 kb EcoRI fragment from C. freundii (Lindberg at al., 1985) was cloned similarly in pAmpRC.

(B) The entericidin genes were cloned individually and in tandem behind an IPTG-inducible *tac*-promoter in the pMS119 expression vectors (Strack et al., 1992). The promoter elements (Ptac, RpoS) are marked with arrowheads, and the Rho-independent terminators are marked with hairpin structures. The gene fragments of *ef-P'* and *ecnR'* flank the *ecnA* and *ecnB* genes of the C. freundii series, which are labeled pCfEcnAB, pCfEcnA, and pCfEcnB. The E. coli series are labeled pEcEcnAB, pEcEcnA, and pEcEcnB. The following restriction enzyme sites are also marked: E (EcoRI), Bg (BglII), Sp (SphI), P (PstI), As (AseI), and Pv (PvuII).

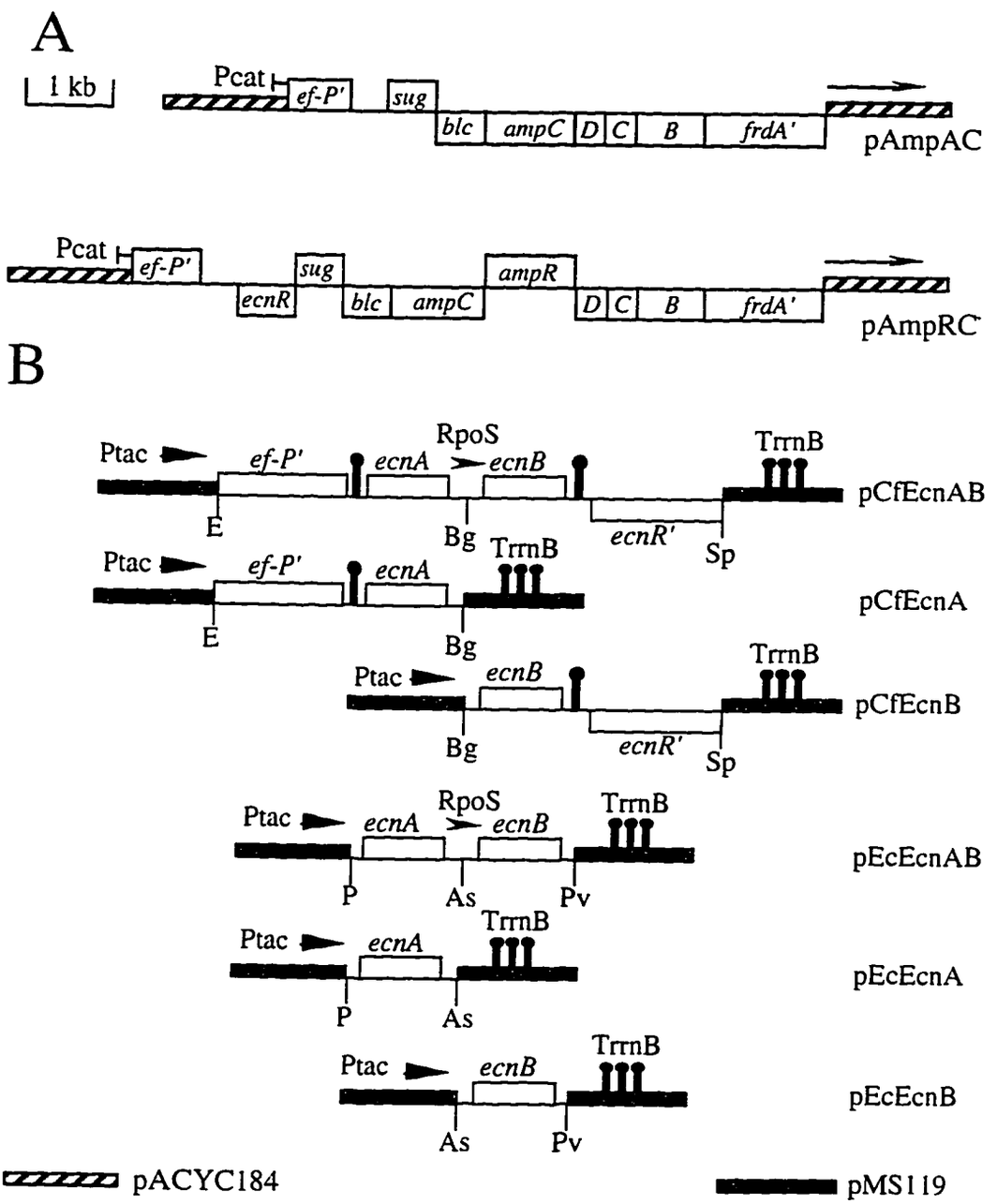

A
1 kb
Pcat
ef-P'
sug
blc
ampC
D
C
B
frdA'
pAmpAC
Pcat
ef-P'
sug
ecnR
blc
ampC
ampR
D
C
B
frdA'
pAmpRC
B
Ptac
ef-P'
ecnA
RpoS
ecnB
TrrnB
E
Bg
ecnR'
Sp
pCfEcnAB
pCfEcnA
pCfEcnB
P
As
Pv
pEcEcnAB
pEcEcnA
pEcEcnB
pACYC184
pMS119

Figure 8-4. Individual and Combined Contribution of the Entericidins to Bacteriolysis. Growth experiments were performed aerobically in Klett flasks at 37C in Luria-Bertani medium supplemented with 300mM NaCl (+ NaCl) or without (- NaCl) and with 1mM IPTG (+ IPTG) or without (- IPTG). E. coli MC4100 was transformed with the C. freundii series of entericidin plasmids and with pMS119EH as a control. pMS119EH (closed squares), pCfEcnAB (open squares), pCfEcnA (closed circles), pCfEcnB (open circles).

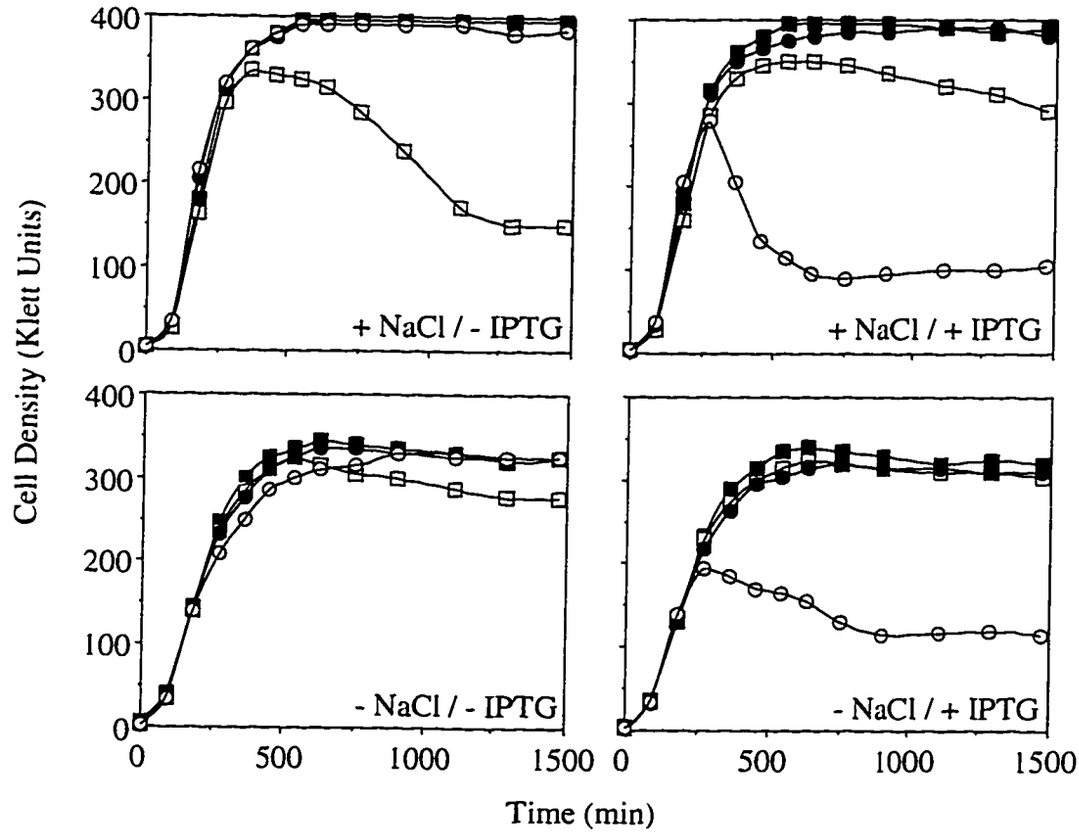

Cell Density (Klett Units)
400
300
200
100
0
+ NaCl / - IPTG
+ NaCl / + IPTG
- NaCl / - IPTG
- NaCl / + IPTG
0
500
1000
1500
Time (min)

Figure 8-5. Palmitate Labeling and Membrane Localization of the Entericidins.

(A) E. coli MC4100 was transformed with pEcEcnA, pEcEcnB, and pMS119HE as a control. Cells were cultured aerobically for 2 hrs in Luria-Bertani medium supplemented with 10 μCi/mL ^{3}H-palmitate in the presence of 1mM IPTG. Protein was resolved with the Tricine-SDS-PAGE system (Schägger and von Jagow, 1987) and analyzed by fluorography (Sankaran et al., 1995). The positions of EcnA, EcnB, and the murein lipoprotein (Mlp) are marked with arrowheads.

(B) E. coli MC4100 was transformed with pCfEcnB and pMS119EH as a control. Cells were grown in the presence of 10 μCi/mL ^{3}H-palmitate for 1 hr before adjustment to 1mM in IPTG and 100 μg/mL in globomycin, where indicated (+), for an additional hour. Protein was resolved with the Tricine-SDS-PAGE system and analyzed by fluorography. EcnB, Mlp, and their diacylglyceryl-prolipoprotein derivatives (DAG-Pro-Mlp and DAG-Pro-EcnB) are marked by arrowheads.

(C) E. coli MC1061 was transformed with pEcEcnA, pEcEcnB, pEcEcnAB, and pMS119HE as a control. Tricine-SDS-PAGE gels were run with membranes prepared from induced cells as described (Bishop et al., 1995) and silver stained. The positions of Mlp, EcnB, and EcnA are marked by arrowheads.

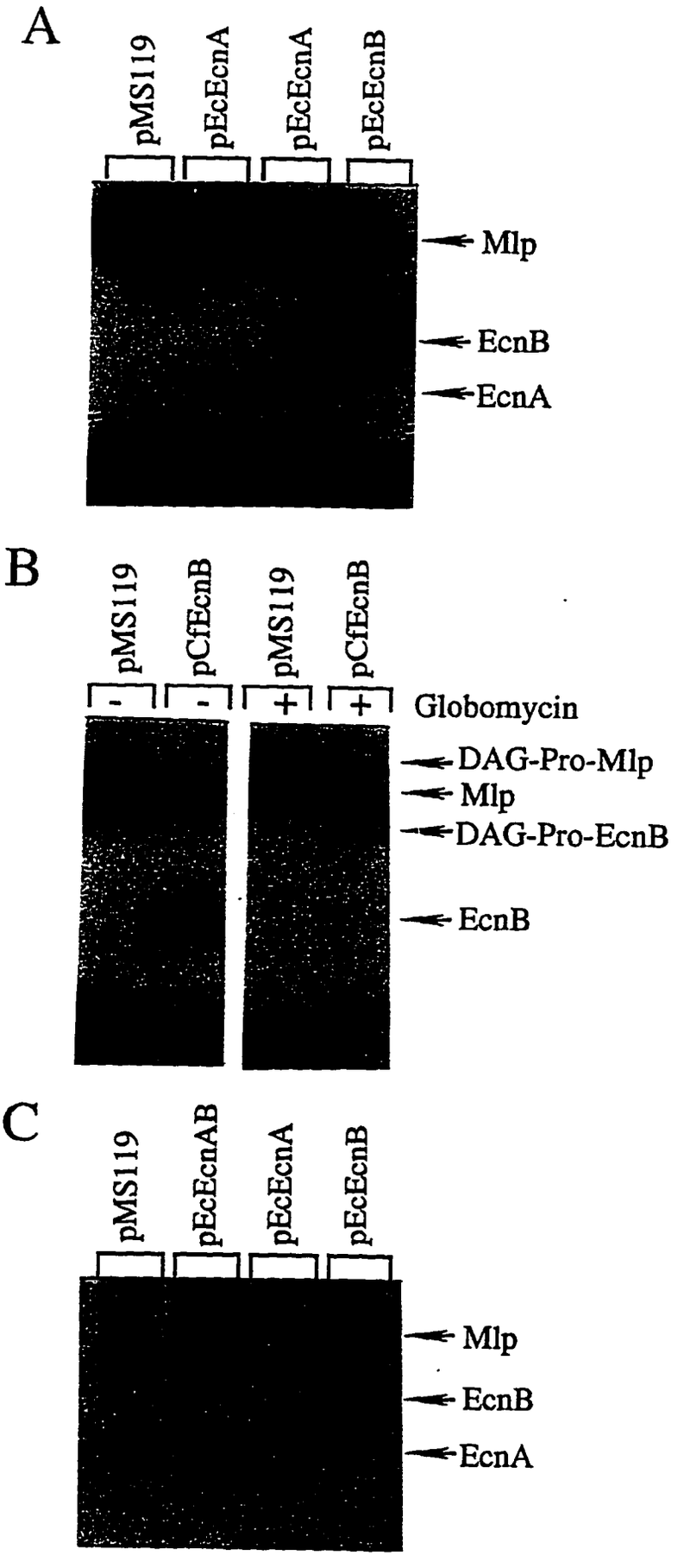

A
pMS119
pEcEcnA
pEcEcnA
pEcEcnB
Mlp
EcnB
EcnA
B
pMS119
pCfEcnB
pMS119
pCfEcnB
-
-
+
+
Globomycin
DAG-Pro-Mlp
Mlp
DAG-Pro-EcnB
EcnB
C
pMS119
pEcEcnAB
pEcEcnA
pEcEcnB
Mlp
EcnB
EcnA

Figure 8-6. Conformational Analysis of the Entericidins.

(A) Helical wheel representations (Schiffer and Edmundson, 1967) of the E. coli entericidins starting from the putative lipoprotein processing sites. Hydrophobic residues are boxed and positive (+) or negative (-) charges are encircled.

(B) Far-ultraviolet circular dichroism spectra of N^{α}-acetylserine derivatives of entericidin A (squares) and entericidin B (circles). Open symbols indicate spectra recorded under benign conditions (50mM potassium phosphate pH 7 containing 50mM KCl at 25 C). Solid symbols indicate spectra recorded from samples supplemented with 50% trifluoroethanol.

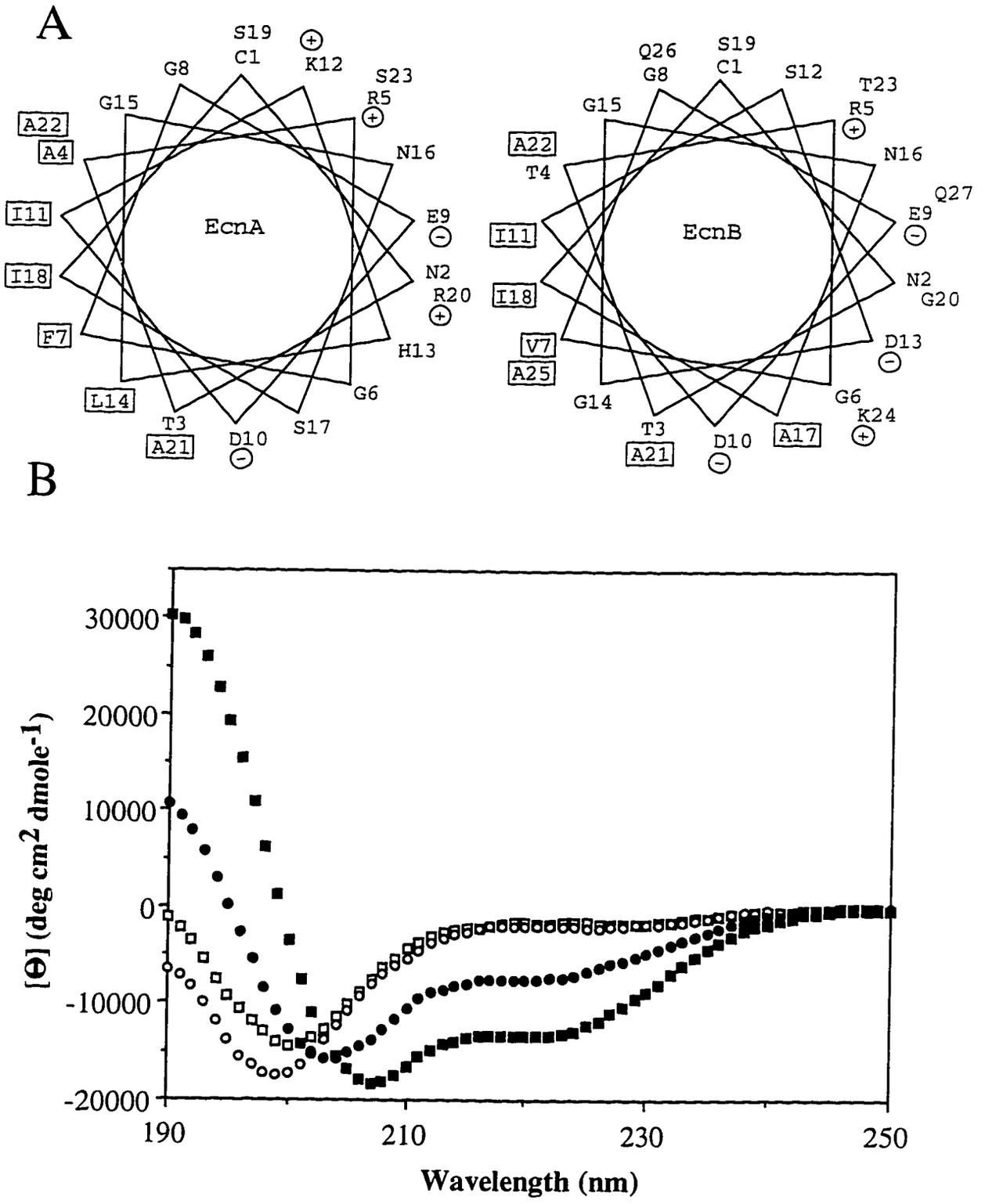
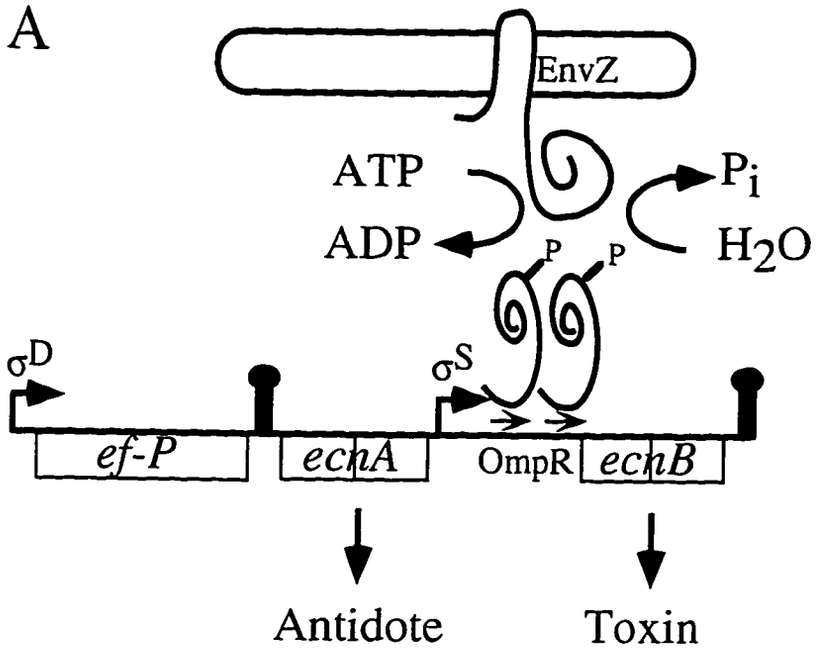
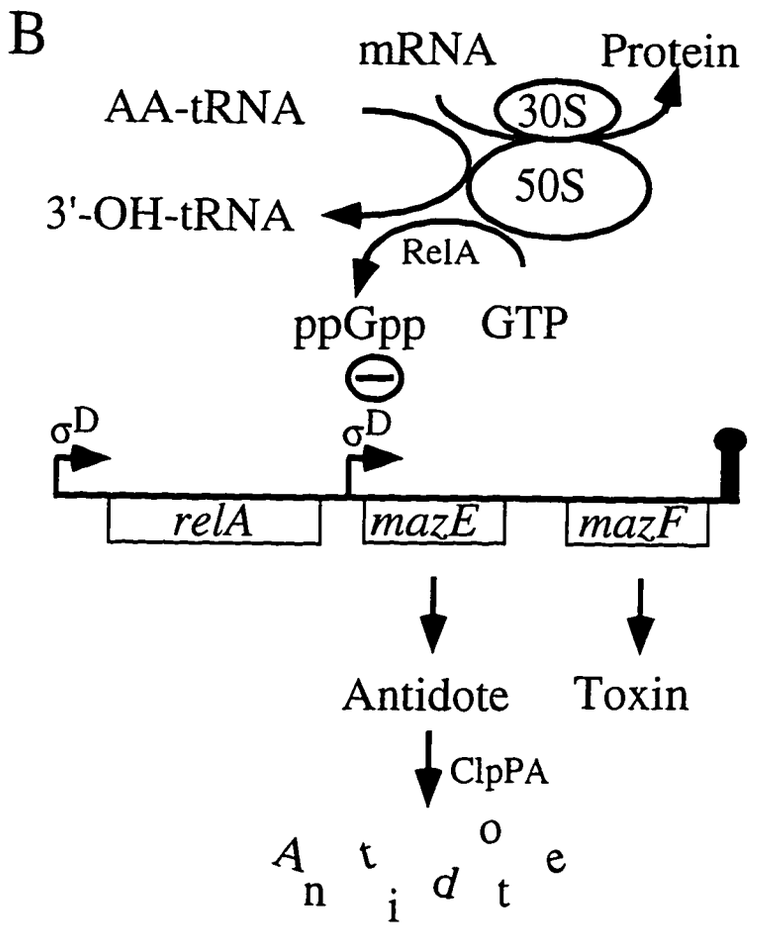

A
S19
C1
K12
G8
S23
R5
G15
A22
A4
N16
I11
EcnA
E9
I18
N2
R20
F7
H13
L14
G6
T3
A21
D10
S17
Q26
G8
S19
C1
S12
T23
R5
G15
A22
T4
N16
Q27
I11
EcnB
E9
I18
N2
G20
V7
D13
A25
G14
G6
T3
A21
D10
A17
K24
B
30000
20000
10000
0
-10000
-20000
190
210
230
250
[Θ] (deg cm2 dmole-1)
Wavelength (nm)

Figure 8-7. Comparison of the Entericidin Locus with the *mazEF* Addiction Module. (A) The entericidin locus of E. coli is composed of the paralogous *ecnA* and *ecnB* genes, which encode an antidote and a bacteriolytic toxin, respectively, and are located downstream of the elongation factor-P operon (*ef-P*). The *ecnB* promoter belongs to a subclass of *rpoS*-dependent promoters that are active only under high osmolarity conditions. Osmotic repression by the EnvZ/OmpR two-component signal transduction pathway ensures that EcnB expression is governed by a stochastic genetic switch. (B) The *mazEF* addiction module is composed of an antidote/toxin gene pair that is located downstream of the *relA* operon. RelA is a ribosomal subunit that senses uncharged aminoacyl-tRNAs and responds by synthesizing guanosine-3',5'-bispyrophosphate (ppGpp), which is an intracellular starvation signal of the stringent response. High levels of ppGpp repress transcription of *mazEF*, which leads to an unmasking of the MazF toxin due to proteolytic degradation of the MazE antidote by the ClpPA protease. Promoters are indicated as σ^D or σ^S and Rho-independent terminators are indicated by hairpin structures.


A
EnvZ
ATP
ADP
Pi
H2O
P
P
σD
σS
ef-P
echA
OmpR
echB
Antidote
Toxin


B
mRNA
Protein
AA-tRNA
30S
50S
3'-OH-tRNA
RelA
ppGpp
GTP
σD
σD
relA
mazE
mazF
Antidote
Toxin
ClpPA
Antidote

References

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996). An *Escherichia coli* chromosomal "addiction module" regulated by guanosine-3',5'-bispyrophosphate: a model for programmed bacterial cell death. Proc. Natl. Acad. Sci. U.S.A. *93*, 6059-6063; published erratum, ibid. *93*, 9991.

Alexander, D.M., and St John, A.C. (1994). Characterization of the carbon starvation-inducible and stationary phase-inducible gene slp encoding an outer membrane lipoprotein in *Escherichia coli*. Mol. Microbiol. *11*, 1059-1071.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Ameisen, J.C. (1996) The origin of programmed cell death. Science *272*, 1278-1279.

Aoki, H., Adams, S.L., Chung, D.G., Yaguchi, M., Chuang, S.E., and Ganoza, M.C. (1991). Cloning, sequencing and overexpression of the gene for prokaryotic factor EF-P involved in peptide bond synthesis. Nucleic Acids Res. *19*, 6215-6220.

Bishop, R.E., and Weiner, J.H. (1993a). Overproduction, solubilization, purification and DNA-binding properties of AmpR from *Citrobacter freundii*. Eur. J. Biochem. *213*, 405-412.

Bishop, R.E., and Weiner, J.H. (1993b). Complementation of growth defect in an *ampC* deletion mutant of *Escherichia coli*. FEMS Microbiol. Lett. *114*, 349-354.

Bishop, R.E., Penfold, S.S., Frost, L.S., Höltje, J-.V., and Weiner, J.H. (1995). Stationary phase expression of a novel *Escherichia coli* outer membrane lipoprotein and its relationship with mammalian apolipoprotein D: implications for the origin of lipocalins. J. Biol. Chem. *270*, 23097-23103.

Boguski, M.S., Birkenmeier, E.H., Elshourbagy, N.A., Taylor, J.M., and Gordon, J.I. (1986). Evolution of the apolipoproteins: structure of the rat apo-A-IV gene and its relationship to the human genes for apo-A-I, C-III, and E. J. Biol. Chem. *261*, 6398-6407.

Burland, V., Plunkett, G., Sofia, H.J., Daniels, D.L., and Blattner, F.R. (1995). Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res. *23*, 2105-2119.

Casadaban, M.J., and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. *138*, 179-207.

Ding, Q., Kusano, S., Villarejo, M., and Ishihama, A. (1995). Promoter selectivity control of *Escherichia coli* RNA polymerase by ionic strength: differential recognition of osmoregulated promoters by $E\sigma^D$ and $E\sigma^S$ holoenzymes. Mol. Microbiol. *16*, 649-656.

Ehlert, K., Höltje, J.V., and Templin, M.F. (1995). Cloning and expression of a murein hydrolase lipoprotein from *Escherichia coli*. Mol. Microbiol. *16*, 761-8.

Epand, R.M., Shai, Y., Segrest, J.P., and Anantharamaiah, G.M. (1995). Mechanisms for the modulation of membrane bilayer properties by amphipathic helical peptides. Biopolymers *37*, 319-338.

Espinosa-Urgel, M., Chamizo, C., and Tormo, A.A. (1996). Consensus structure for σ^S-dependent promoters. Mol. Microbiol. *21*, 657-659.

Greener, T., Govezensky, D., and Zamir, A. (1993). A novel multicopy suppressor of a *groEL* mutation includes two nested open reading frames transcribed from different promoters. EMBO J. *12*, 889-896.

Greenfield, N., and Fasman, G.D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry *8*, 4108-4116.

Harlocker, S.L., Bergstrom, L., and Inouye, M. (1995). Tandem binding of six OmpR proteins to the *ompF* upstream regulatory sequence of *Escherichia coli*. J. Biol. Chem. *270*, 26849-26856.

Hengge-Aronis, R., Klein, W., Lange, R., Rimmele, M., and Boos, W. (1991). Trehalose synthesis genes are controlled by the putative sigma factor encoded by *rpoS* and are involved in stationary-phase thermotolerance in *Escherichia coli*. J. Bacteriol. *173*, 7918-7924.

Hengge-Aronis, R. (1996a). Back to log phase: σ^S as a global regulator in the osmotic control of gene expression in Escherichia coli. Mol. Microbiol. *21*, 887-893.

Hengge-Aronis, R. (1996b). Regulation of gene expression during entry into stationary phase. In *Escherichia coli* and *Salmonella*, 2nd ed., F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington D.C.: ASM Press), pp. 1497-1512.

Höltje, J.-V. (1995). From growth to autolysis: the murein hydrolases in *Escherichia coli*. Arch. Microbiol. *164*, 243-254.

Hsieh, S.-Y., Ko, T.-P., Tseng, M.-Y., Ku, W.-Y, Chak, K.-F., and Yuan, H.S. (1997). A novel role of ImmE7 in the autoregulatory expression of the ColE7 operon and identification of possible RNase active sites in the crystal structure of dimeric ImmE7. EMBO J. *16*, 1444-1454.

Hussain, M., Ichihara, S., and Mizushima, S. (1980). Accumulation of glyceride-containing precursor of the outer membrane lipoprotein in the cytoplasmic membrane of *Escherichia coli* treated with globomycin. J. Biol. Chem. *255*, 3707-3712.

Jacobs, C., Frère, J.M., and Normark, S. (1997). Cytosolic intermediates for cell wall biosynthesis and degradation control inducible beta-lactam resistance in gram-negative bacteria. Cell *88*, 823-832.

Jensen, R.B., and Gerdes, K. (1995). Programmed cell death in bacteria: proteic plasmid stabilization systems. Mol. Microbiol. *17*, 205-10.

Jung, J.U., Gutierrez, C., Martin, F., Ardourel, M., and Villarejo, M. (1990). Transcription of *osmB*, a gene encoding an *Escherichia coli* lipoprotein, is regulated by dual signals: osmotic stress and stationary phase. J. Biol. Chem. *265*, 10574-10581.

Lavigne, P., Kondejewski, L.H., Houston, M.E.,Jr., Sönnichsen, F.D., Lix, B., Sykes, B.D., Hodges, R.S., and Kay, C.M. (1995). Preferential heterodimeric parallel coiled-coil formation by synthetic max and c-myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization. J. Mol. Biol. *254*, 505-520.

Lecuc, M., Ishidate, K., Shakibai, N., and Rothfield, L. (1992). Interactions of *Escherichia coli* membrane lipoproteins with the murein sacculus. J. Bacteriol. *174*, 7982-7988.

Lindberg, F., Westman, L., and Normark, S. (1985). Regulatory components in *Citrobacter freundii ampC* beta-lactamase induction. Proc. Natl. Acad. Sci. U.S.A. *82*, 4620-4624.

Macquaire, F., Baleux, F., Giaccobi, E., Huynh-Dinh, T., Neumann, J.-M., and Sanson, A. (1992). Peptide secondary structure induced by a micellar phospholipidic interface: proton nmr conformational study of a lipopeptide. Biochemistry *31*, 2576-2582.

Muchmore, S.W., Sattler, M., Liang, H., Meadows, R.P., Harlan, J.E., Yoon, H.S., Nettesheim, D., Chang, B.S., Thompson, C.B., Wong, S.L., Ng, S.L., and Fesik, S.W. (1996). X-ray and NMR structure of human Bcl-x_L, an inhibitor of programmed cell death. Nature *381*, 335-341.

Nikaido, H. (1994). Prevention of drug access to bacterial targets: permeability barriers and active efflux. Science *264*, 382-388.

Nikaido, H. (1996). Outer membrane. In *Escherichia coli* and *Salmonella*, 2nd ed., F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington D.C.: ASM Press), pp. 29-47.

Park, J.T. (1996). The murein sacculus. In *Escherichia coli* and *Salmonella*, 2nd ed., F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington D.C.: ASM Press), pp. 48-57.

Paulsen, I.T., Skurray, R.A., Tam, R., Saier, M.H.,Jr, Turner, R.J., Weiner, J.H., Goldberg, J.H., and Grinius, L.L. (1996). The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. Mol. Microbiol. *19*, 1167-1175.

Pratt, L.A., and Silhavy, T.J. (1994). OmpR mutants specifically defective for transcriptional activation. J. Mol. Biol. *243*, 579-594.

Pratt, L.A., and Silhavy, T.J. (1995). Identification of base pairs important for OmpR-DNA interaction. Mol. Microbiol. *17*, 565-573.

Pugsley, A.P. (1993). The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. *57*, 50-108.

Pugsley, A.P,. and Schwartz, M. (1984). Colicin E2 release: lysis, leakage or secretion? Possible role of a phospholipase. EMBO J. *3*, 2393-2397.

Rosenbush, J.P. (1974). Characterization of the major envelope protein from *Escherichia coli*: regular arrangement on the peptidoglycan and unusual dodecyl sulfate binding. J. Biol. Chem. *249*, 8019-8029.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sankaran, K., Gupta, S.D., and Wu, H.C. (1995). Modification of bacterial lipoproteins. Meth. Enzymol. *250*, 683-697.

Schägger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. *166*, 368-379.

Schiffer, J, and Edmundson, A.B. (1967). Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. *7*, 121-135.

Sönnichsen, F.D., van Eyk, J.E., Hodges, R.S., and Sykes, B.D. (1992). Effect of trifluoroethanol on protein secondary structure: an nmr and cd study using a synthetic actin peptide. Biochemistry *31*, 8790-8798.

Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992). A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the alpha protein of the *Escherichia coli* satellite phage P4. J. Biol. Chem. *267*, 13062-13072.

Tytler, E.M., Segrest, J.P., Epand, R.M., Nie, S.Q., Epand, R.F., Mishra, V.K., Venkatachalapathi, Y.V., and Anantharamaiah, G.M. (1993). Reciprocal effects of apolipoprotein and lytic peptide analogs on membranes: cross-sectional molecular shapes of amphipathic alpha helixes control membrane stability. J. Biol. Chem. *268*, 22112-22118.

van der Wal, F.J., Luirink, J., and Oudega, B. (1995). Bacteriocin release proteins: mode of action, structure, and biotechnological application. FEMS Microbiol. Rev. *17*, 381-399.

Williams, J.G., and Mason, P.J. (1985). In Nucleic Acids Hybridization: a Practical Approach, B.D. Hames and S.J. Higgens eds., (Oxford, UK: IRL Press), pp. 139-178.

Wishart, D.S., Boyko, R.F., Willard, L., Richards, F.M., and Sykes, B.D. (1994). SEQSEE: a comprehensive suite for protein sequence analysis. CABIOS *10*, 121-132.

Wu, H.C. (1996). Biosynthesis of lipoproteins. In *Escherichia coli* and *Salmonella*, 2nd ed., F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, eds. (Washington D.C.: ASM Press), pp. 1005-1014.

Yamaguchi, K., Yu, F., and Inouye, M. (1988). A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. Cell *53*, 423-32.

Yarmolinsky, M.B. (1995). Programmed cell death in bacterial populations. Science *267*, 836-837.

Zagorski, M.G., Norman, D.G., Barrow, C.J., Iwashita, T., Tachibana, K., and Patel, D.J. (1991). Solution structure of pardaxin P2. Biochemistry *30*, 8009-80017.

Zambrano, M.M., and Kolter, R. (1996). Gasping for life in stationary phase. Cell *86*, 181-184.

Zambrano, M.M., Siegele, D.A., Almirón, M., Tormo, A., and Kolter, R. (1993). Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. Science *259*, 1757-1760.

CHAPTER 9

DISCUSSION

Perhaps the most striking feature of the *frd-efp* chromosome interval of enteric bacteria is that the component genes contribute either directly to general bacterial resistance or are implicated by their evolutionary design. Several components are related to resistance determinants associated with bacterial plasmids, which revealed comparisons and contrasts between chromosomal and plasmid-borne resistance determinants. While it is clear that resistance genes can be mobilized between plasmids and chromosomes, it is also clear that natural selection acts differently on related resistance determinants when they are in plasmid or chromosomal contexts. These differences can be understood in terms of the subtler aspects of bacterial genetics if bacterial genetics is integrated into modern evolutionary theory. What follows is a modest attempt to make this integration in order to understand the adaptive value of the *frd-efp* interval.

The neo-Darwinian synthesis

The theory of evolution is the most important and most widely misunderstood concept in biology. Attributed to Charles Darwin and A. R. Wallace, the theory of evolution as stated in Darwin's "On the Origin of Species by Means of Natural Selection" (1) was essentially correct, but wrong where it accepted the inheritance of acquired characters. Darwin's formulation was based on the "individual-selectionist" paradigm; that is, that natural selection acts on phenotypic differences between individual organisms within a population. The population genetic theories of the 1920's and 1930's were the basis of the neo-Darwinian synthesis, which rectified the theory of evolution by rejecting the inheritance of acquired characters, and integrating Mendelian genetics into evolutionary theory (2). However, the neo-Darwinian synthesis was followed with

confusing ideas of how selection can act at levels above the individual, such as the selection of groups of related individuals, populations, species, and communities. This led to an advocacy of the "principle of parsimony" in evolution (3) - if individual selection can explain adaptations, higher levels of selection should not be invoked. A trend of holistic thinking in evolutionary biology is voiced by Mayr (4, p. 588): "naturalists from Darwin on and the more perceptive geneticists have always emphasized that not genes but whole organisms - potentially reproducing individuals - are the unit of selection."

However, the individual selectionist paradigm of neo-Darwinism developed simultaneously with that of the reductionist central dogma of molecular biology. Accounting for the inherent degeneracy of the genetic code, Kimura (5) pronounced the neutral theory of molecular evolution, which sparked the 'neutralist-selectionist controversy'. Kimura merely drew attention to the contrasting rates and modes of evolution at the phenotypic and molecular levels (6): "Here it is important to note that natural selection acts directly on phenotype, but only secondarily on the molecular constitution of genes." Therefore (6): "The theory does not deny the role of natural selection in determining the course of adaptive evolution, but it assumes that only a minute fraction of DNA changes are adaptive in nature." While molecular biology demonstrates that acquired characters cannot be inherited, it also demonstrates that nucleic acid is confined *within* individuals, or rather, within their cells. Therefore, the most recent shift in evolutionary thought extends the principle of parsimony to account for the fact that nucleic acid operates below the level of the individual organism. As noted by Medawar (7): "The most important single innovation in the modern synthesis was however the new conception that a population that was deemed to undergo evolution could best be thought of as a population of fundamental replicating units - of genes - rather than as a population of individual animals or of cells." This modern paradigm of "gene-selectionism" can be regarded as a variation on the theme of individual

selectionism, but it is a more reliable and powerful variation. Initial objections to gene selectionism are usually directed at the recognition of the "gene" as the unit of selection, but this is easily clarified if one distinguishes units of selection from units of function.

Perhaps the clearest reductionist separation of units of selection and function has been achieved by Dawkins' concept of replicators and vehicles (8). These can be illustrated by drawing an analogy with the laws of thermodynamics. The first law of evolution states that *replicators are driven by vehicles.* A replicator is a substance that makes copies of itself and obeys the laws of quantum mechanics, where a vehicle is an energy transducer that obeys the laws of thermodynamics. The two are inseparable, because replication is an endothermic process. The second law of evolution states that *mutable replicators experience evolution.* Evolution, or "the differential survival of alternative replicators" (8, p.82), is inevitable when a mutable replicator is driven by its vehicle. Evolution accounts for a spatial component (diversification of replicators), the mechanism of which is mutation, and a temporal component (transformation of vehicles), the mechanism of which is selection. Although these abstract laws can apply to any objects, organic evolution is based on replicators composed of nucleic acid, which resides within vehicles known as cells. Gene selectionism is so named because it recognizes that the unit of selection - the replicator - can be, at its most elementary level, equivalent to a gene. But higher order replicators can be regarded as substantially more than genes.

Gene selectionism, or more generally, replicator selectionism, recognizes that natural selection acts on the phenotype, which is the environmentally influenced outward manifestation of genes. But the phenotype is only the means by which genes lever themselves into the next generation. Only the genetic replicator is the unit of selection, because only the replicator is inherited. As Dawkins notes (8, p. 114), the phenotype "is the all important instrument of replicator preservation: it is *not* that which is preserved." Therefore, it is correct to say that the teleonomy of a genetic replicator is that it is an adaptation for its own survival. Replicator survival may be achieved if the replicator

contributes to the survival of the vehicle in which it happens to reside. The benefit to the individual vehicle may support an individual selectionist interpretation, but this is conditional and best viewed as secondary to the primary consequence - survival of the replicator. As noted by Dawkins (8, p. 114):

> The obvious and archetypal vehicle is the individual organism, but this may not be the only level in the hierarchy of life at which the title is applicable. We can examine as candidate vehicles chromosomes and cells below the organism level, groups and communities above it. At any level, if a vehicle is destroyed, all the replicators inside it will be destroyed. Natural selection will therefore, at least to some extent, favour replicators that cause their vehicles to resist being destroyed.

This conceptual reorientation from individual to gene selectionism provides an additional perspective that can explain several longstanding paradoxes in microbiology. Unfortunately, the most eloquent advocacy of gene selectionism (8) focuses primarily on Mendelian systems and makes only passing references to bacteria. In light of the current renaissance in microbial biology (9), it is worthwhile to consider the gene-selectionist paradigm from a bacterial viewpoint.

Replicator-selectionism in bacteria

Dawkin's formulation of gene-selectionism hinges on the elasticity (8, p. 90) of the "gene" definition as applied to Mendelian populations. Whether bacterial or Mendelian, genes are units of both selection and function. Based on the *cis-trans* complementation test, Benzer (10) coined the terms muton, recon, and cistron to account for the elementary functional units of heredity. Traditionally, a group of cistrons that specify a polypeptide chain were regarded fundamentally as a gene, but the discovery of ribozymes indicates that it is necessary to account for the specification of any RNA transcript. Jacob and other molecular biologists have coined terms for several higher order functional units of heredity, ranging from the operon to the integron (11). Since

"integron" has been recently recoined by molecular biologists (who failed to account for Jacob's earlier usage of the term) to describe transposons that carry multiple antibiotic resistance (12), perhaps we can forgive Dawkins for not recognizing that Jacob also coined the term "replicator" to describe "an operator of DNA replication" (13). As for the "gene" as a unit of selection, Dawkins has appropriately utilized the definition proposed by G.C. Williams (3) as "that which segregates and recombines with appreciable frequency." This definition works in Mendelian populations because the fragmenting effects of meiotic recombination necessitates the rejection of the chromosome as the unit of selection. Recombination randomizes the association between alleles at different loci and reduces any linkage disequilibrium in the population (14, p. 555). As noted by Dawkins (8, p. 91): "An active replicator is a chunk of genome that, when compared to its alleles, exerts phenotypic power over its world, such that its frequency increases or decreases relative to its alleles." In this spirit, the unit of selection - the replicator - can be regarded at its most elementary level in the same way as a group of cistrons that specify an RNA transcript, and so it happens that replicators and genes can be the same thing. But this is conditional on the assumption that recombination has reduced any linkage disequilibrium in the population.

Maynard-Smith notes that (15) "The population genetics of bacteria is difficult because recombination is neither absent nor very common." In fact, the manifestation of low recombination rates in bacteria is strong linkage disequilibrium. According to Slatkin (16), "It is clear that when permanent linkage disequilibrium is maintained in a population, the higher order interactions are important and the chromosome tends to act as a unit. The degree to which this is true in any given system is a measure of whether the gene or the chromosome is the unit of selection, or, more accurately, what parts of the genome can be said to be in unison," Before we recognize the chromosome as the unit of selection in bacteria, we must first recognize that linkage disequilibrium is only strong and not permanent, a fact that obscures the concept of a bacterial species.

One effort to rectify this ambiguity regards the "meroclone" concept of Milkman (17). Bacteria are essentially clonal in nature and the diversification of the bacterial chromosome is limited by a conservative process known as periodic selection (18). As spontaneous mutants arise in populations, they comprise a numerical minority in the larger pool of the parental cell line. In the absence of selection, one would predict that the population would gradually degrade into a collective of mutant lines. However, populations tend to remain unchanged over long periods in the absence of selection. This results from an unexplained phenomenon where specific mutants arise periodically, within a given number of generations, depending on the growth conditions, and replace all mutant lines that have arisen since the previous replacement event. The mutant that replaces the population derives from the line of cells which are numerically superior and usually represents the parental line. The result is that a prototrophic bacterium will remain prototrophic in the absence of selection for prototrophy. Thus, periodic selection leads to the persistence, within a population, of uniformity in sizable genomic regions of the bacterial chromosome. A true clonal population will exhibit uniformity in the entire chromosome. However, if selection acts on a mutant line so that it becomes the numerically superior component of another population, it will be established by the next replacement event and will be maintained in that population. This may lead to the incorporation of adaptive genes and the elimination of maladaptive genes. Therefore, low rates of recombination will eventually introduce small interruptions in clonal frames and establish a meroclone. According to Milkman (17), "A meroclone is thus a group of entities in each of which a majority of the genetic makeup is derived from a given most recent common ancestor." In the natural environment, clonal frames will diverge to the point that we will recognize distinct bacterial species. Eventually, over extended periods, clonal frames become obliterated with increasing evolutionary distance. In the end, the elementary unit of selection - the replicator - can be regarded as the gene. But strong linkage disequilibrium necessitates that replicators compared between closely related

species can comprise many genes, and within an individual bacterium, unlike Mendelian systems, the entire chromosome can be effectively regarded as a single replicator.

If the unit of selection - the replicator - is a recombinational unit, then we can distinguish replicators within genomes. In this case, the key to identifying the target of selection is to identify those components within the genome that recombine with similar frequencies, because these genetic elements will be selected independently. To recognize why this is so, we must first clarify a misconception regarding the role of recombination in bacteria. Levin and Lenski (19, p. 123) state:

> Save for the incorporation of free DNA (a process known as transformation), plasmids and phage are the sole vectors for the exchange of genetic material between bacteria. Thus, it might seem that a good deal of the coevolution of these organisms would be directed toward the role of plasmids and phage as vehicles of recombination. We suggest that this has not been the case. Although gene exchange mediated by these vectors is unquestionably important to bacterial adaptation and evolution, it is unlikely that natural selection has acted directly to increase the effectiveness of these replicons as vectors for recombination.

In other words, recombination probably originated in bacteria for the interspecific acquisition, by horizontal transfer, of useful genetic adaptations. A possible consequence of recombination is that parasitic DNAs were engendered as a byproduct.

The phenomenon of non-phenotypic selection was recognized by Doolittle and Sapienza (20), who noted that "natural selection operating within genomes will inevitably result in the appearance of DNAs with no phenotypic expression whose only 'function' is survival within genomes." While these parasitic DNAs can be produced and discarded by recombination, it is inevitable that some elements are harder to eliminate than to create. This may have started with a DNA element that over-replicates the host and avoids elimination by recombining into random positions in the genome. In this respect, an

insertion sequence is probably the simplest known parasitic DNA. If an insertion sequence acquires the ability to control its own transposition it can become a transposon. A plasmid might be engendered by the premature excision of a replication origin. If a plasmid acquires the means for its own transmission it can become a conjugative plasmid or a virus. In the process of acquiring the means to control their own dissemination, the source of genetic adaptations for parasitic DNA is euchromosomal DNA. Whatever the origin of parasitic DNAs, their adaptive evolution is unique from their chromosomal counterparts to the extent that they are not continually subjected to the same selection acting on the chromosome.

As noted by Campbell (21, p. 65), "If existing elements frequently have an ancient pedigree as accessory elements, we may ask not just how they are maintained at this moment, but what are all the factors that have allowed them, and not their competitors, to come down to us from antiquity." Since the host cell does not gain any essential functions from parasitic DNAs, they are a metabolic burden to the cell. Campbell (21, p. 63) continues: "Although an element with a negative effect on the host can be maintained by over-replication, long-term selection in competition with other hosts should ultimately eliminate it. Thus, long-term survival may require that the element earn its keep; i.e. whatever cost is entailed in perpetuating the element should at least be balanced by some positive contribution to the organism's phenotype." Perhaps a consequence is that parasitic DNAs are presently adapted to contribute to the diversification of the genome by promoting DNA rearrangements that were not previously possible. But this cannot be the driving force behind their evolution, because each step-by-step Darwinian adaptation must be favorable in itself; surely, genomic rearrangements emerged only after parasitic DNAs had begun to flourish (22). Another way that a parasitic DNA can earn its keep is to harbor genes that endow a "useful" phenotype on the host. Campbell (21) has argued that the cell should be expected to extract such useful genes and then eliminate the remainder of the parasitic element. Therefore, we can expect that the specific properties

of genes that present day parasitic DNAs utilize to endow useful phenotypes on their hosts will indicate how those genes successfully avoided extraction.

Why does natural selection distinguish between different replicators, within bacterial genomes, on the basis of their relative frequencies of recombination? The answer lies in the fact that recombination frequencies reflect the tendency of replicators to exploit horizontal transmission. As noted by Dawkins (8, p. 224), "an important message is that replicating entities will tend to work against each other to the extent that they employ different methods of egress from vehicle to vehicle." In bacteria, all genes lever themselves into the next vehicle by vertical transfer, but they differ with respect to their ability to lever themselves into an entirely different vehicle by horizontal transfer. Since natural selection acts on replicators in the context of the vehicles they inhabit, replicators will evolve independently in different vehicles. Therefore, we must recognize that the chromosome has evolved under periodic selection for vertical transmission and is primarily limited to the vehicle it encodes. Parasitic DNAs have overcome this limitation and their evolutionary history reflects the 'commonwealth' of organisms they have inhabited. Levin and Lenski (19) comment on the importance of the density of habitable cells in populations, and of structured and non-structured habitats, in the horizontal transmission of parasitic DNAs. We can expect that parasitic DNAs will utilize mechanisms to sense cell density and habitat so as to appropriately execute their programs of vertical transfer or, when superior vehicles are accessible, to activate programs of horizontal transfer.

In summary, the important distinction between euchromosomal and parasitic genes is the investment in the vehicle of transmission. The survival of a cell hinges on a balance between replication of the genome and appropriate management of cellular metabolism, which may require that the cell enter into transient periods of dormancy. The parasitic DNA survives preferably by inhabiting a vehicle that is replicating or that can be convinced to do so. If the vehicle should enter into dormancy, the parasitic DNA

is adapted to seek out another vehicle that continues to be concerned with replication. Should no such replicating vehicle be available, the only option of the parasitic DNA will be to resign itself to a state of dormancy. In this regard, dormancy is an important signal for any parasitic DNA.

Teleonomic significance of antibiotic resistance and virulence

Parasitic DNAs differ in terms of their ability to govern their own dissemination. While phage and conjugative plasmids encode the machinery for their own transmission, non-conjugative plasmids are dependent on transformation and on "hitch-hiking" with conjugative plasmids. Transposons can be transmitted horizontally only if they are integrated into a transmissible element. In addition to coding (or borrowing) their own dissemination functions, parasitic DNAs can acquire mechanisms to earn their keep (21). As stated previously, maintenance can involve the acquisition of a useful metabolic function for the cell. This has been termed "niceness" by Levin and Lenski (19). However, parasitic DNAs can also exhibit "ruthlessness" when they create the necessity for their own maintenance by encoding a toxin and a corresponding antidote, which is, thereby, imposed on the cell as a useful metabolic function. The dependence of a cell on the parasitic DNA for the antidote of the corresponding toxin is known as "addiction" and is common among plasmids.

Addiction toxins can act directly on intracellular targets or they can be secreted beforehand, in which case they can also function as antibiotics by killing unrelated cells that happen to be sensitive to the toxin. The antibiotic can contribute to the elimination of competitors and also to the provision of nutrients if the killed cells release their cellular contents by lysis. DNA released by lysis can also serve to transform any competent cells in the population (23). Therefore, the overall effect of an antibiotic can depend on whether it is "bacteriostatic" or "bacteriolytic." However, the plasmid will still be maintained in the population to the extent that the antibiotic kills the cells that have lost

the antidote. This will depend in part on the proportion of the population that bears the plasmid.

Bacteriocins (microcins and colicins) were once regarded strictly as peculiar narrow spectrum antibiotics, effective primarily against those cell-types that are compatible with the bacteriocin plasmid (24). However, the narrow spectrum of bacteriocins is easily rationalized with the fact that they are co-expressed with their cognate immunity determinants, both of which are now widely regarded as products of an addiction module (25). A similar situation occurs with the restriction-modification systems. While restriction enzymes can clearly protect cells from foreign DNA, the coupled synthesis of plasmid-borne restriction and modification enzymes is also regarded as an addiction mechanism (22). This has resolved the apparent paradox of the so-called "rare-cutters", which target sites in the chromosome that are unlikely to exist in small foreign DNAs (26).

Antibiotics, specifically, are compounds produced by microorganisms that are harmful to other microorganisms. Antibiotics are a subclass of allelopathic substances that are exchanged between plants and microorganisms in nature (27). Most antibiotics belong to a larger class of microbial compounds known as secondary metabolites (28). The function and origin of secondary metabolites remains one of the longstanding unsolved mysteries of microbiology. Numerous functions of secondary metabolites, especially antibiotics, have been attributed to various aspects of natural selection, but the roles of most secondary metabolites are unknown (29). One proposal argues that secondary metabolites are the remnants of primordial biochemical effectors, some of which have been adapted by natural selection to function as contemporary antibiotics (30). I prefer to think that we just don't understand the evolutionary importance of most secondary metabolites. As stated by Demain (28), the "analysis of the problem of function would probably proceed faster if we stopped attempting to formulate a grand and glorious proposal to explain the function of all antibiotics." However, most explanations

of antibiotics are presently based on the individual selectionist paradigm (29). Indeed, antibiotics can be widely regarded to provide a selective advantage to the producing organism, and perhaps the organism is a true beneficiary when the antibiotics are chromosomally encoded, but many antibiotics are encoded by parasitic DNA. In this case, antibiotics are more easily understood if they are considered to serve adaptive functions for the parasitic DNA that encodes them.

Perhaps the most salient observation of secondary metabolites is that they are primarily produced in stationary phase, often in association with sporulation, and are potentially adaptations for dormancy. Additionally, antibiotic producers tend to be resistant to their own antibiotics when they are producing them, but are sensitive to those same antibiotics when they are provided exogenously in exponential phase (28). This indicates that resistance and production are coupled, which is entirely consistent with the possibility that some antibiotics are simply part of an addiction system, much like the bacteriocins. A logical prediction is that secondary metabolite producers may harbor considerable parasitic DNA. In fact, large segments of the genomes of *Streptomyces*, which express a complex secondary metabolism, are unstable and mobilizable, suggesting that parasitic DNAs are widespread in this genus (31). Truncations of the genetic element that encodes the metabolic pathway for ß-lactam antibiotics (32), which originated in *Streptomyces*, were horizontally transferred to the eukaryotic molds *Penicillium* and *Cephalosporium* (33). These organisms clearly utilize the antibiotics to engage in the microbial arms race, and it is equally clear that many organisms have responded by acquiring resistance determinants as a means of self-defense. This acquired resistance is termed "extrinsic" to distinguish it from "intrinsic" resistance, which results from an innate insensitivity to an antibiotic. If an organism only exhibits intrinsic resistance in stationary phase, this can be known as "phenotypic tolerance" (34). Antibiotics and their resistance determinants are clearly not restricted to addiction

systems, but these may provide a plausible explanation for antibiotics encoded by parasitic DNAs.

While the biological significance of genomic plasticity in *Streptomyces* is still debatable, large chromosomal elements responsible for pathogenesis in *E. coli* are clearly mobilizable parasitic DNAs. Known as pathogenicity islands, these can become elaborate, albeit transient, components of the chromosome, many as large as 200 kilobase pairs. Pathogenicity islands can be recognized by the presence of a multitude of virulence genes, different G+C content from the rest of the chromosome, the presence of mobility determinants, and an inherent instability (35). The pathogenesis that results can be regarded as an extended phenotypic adaptation for the pathogenicity island, which over-rides a dormancy adaptation of the commensal bacterium.

Accumulating evidence supports the hypothesis that most genes associated with parasitic DNAs probably originated in a chromosomal context and were acquired by a process known as "gene pickup." This is illustrated by the serine ß-lactamases, which are homologs of the D-alanyl carboxypeptidase family of peptidoglycan hydrolases (36). *E. coli* harbors a chromosomal serine ß-lactamase, which is optimally expressed at high growth rates, but is insufficiently expressed to provide any significant clinical resistance to ß-lactam antibiotics (37). However, a ß-lactam resistant strain of *E. coli*, isolated from a hospital in Greece, was found to carry a plasmid that encodes another serine ß-lactamase called Rtem (38). This enzyme was so named because the patient was a young girl named Temoniera and the resistance plasmid is known as an R-factor. It is now known that the gene for Rtem resides within a transposable element known as Tn*A*, which was integrated into the R-factor (39). While both Rtem and AmpC clearly originated from D-alanyl carboxypeptidases, only Rtem is normally expressed at levels that provide clinical resistance to ß-lactam antibiotics. Therefore, virulence genes of parasitic DNAs can originate from relatively benign chromosomal sources.

This example raises the question of why a transposon would bother to carry a lone resistance determinant? Unlike the ruthless addiction systems, a lone resistance determinant can only provide "niceness" for the element to earn its keep. But why do transposons primarily carry accessory genes strongly biased toward detoxification functions? Wouldn't any useful metabolic function suffice? As noted by Watanabe (40), "It is difficult to understand why R factors do not carry bacterial markers other than drug resistance, if R factors really did develop through gene pickup. However, this may perhaps be due merely to the fact that the selective pressure for drug-resistant bacteria has been acting strongly in the environment." *Streptomyces* do secrete antibiotics into the soil, which suggests that resistance genes may be selected strongly in this habitat (29). Presumably, antibiotics have been prevalent in the natural environment before they were introduced into clinical practice after the second world war. Indeed, drug resistance determinants exist in antibiotic virgin communities and in bacteria that were isolated before the antibiotic era (44).

Additionally, transposable elements depend on plasmids and phage for their horizontal transmission. Studies have shown that insertion sequences and transposons of *E. coli* are occasionally integrated into the chromosome, but they tend to be preferentially associated with plasmids and phage (22). In accordance with the gene-pickup model, we should expect that the plasmid and phage gene pool is the primary source available to transposons, while euchromosomal genes are only a secondary source. The preponderance of detoxification genes carried by transposons may, to some extent, reflect the tendency of plasmids to harbor genes for resistance to allelopathic substances.

Why, then, don't transposons simply acquire a small intact addiction system? The answer may lie in the specificity of addiction systems for a particular cell type. Non-conjugative plasmids and transposons are highly promiscuous and might require more general mechanisms of earning their keep. Detoxification systems probably provide a general niceness because they allow cells to take over habitats not previously available to

them and for which they have no previous adaptations. However, as noted by Campbell (21), if selection should "favor the retention of the gene and the loss of the rest of the element," then parasitic DNA will contribute to speciation and new habitats for niceness will be depleted. Perhaps hosts are more likely to retain the means to metabolize lactose than, say, resist heavy metals, when these are presented by a parasitic DNA. As noted by Levin (42), "Characters that confer a survival or replication advantage on the individual organisms that express them at a given time or in a given habitat will be favored and evolve at that time in that habitat. Whether the expression of those temporally or locally favored characters will increase or reduce the fitness of that organism at other times or in other habitats is irrelevant." In other words, detoxification genes may be simultaneously favored by the short-sighted evolution of the parasitic DNA and disfavored by periodic selection acting over the long term evolution of the host. Detoxification genes may, therefore, be present in modern transposons and R-factors precisely because they don't persist among potential hosts.

While chromosomal genes can be exploited by parasitic DNAs, it is clear from the above considerations that the reverse is also possible if selection doesn't act against it. Selection may act against the retention of detoxification and antibiotic resistance genes if they facilitate the colonization of habitats where the organism is unlikely to persist over the long term. However, if the gene allows an organism to resist occasional toxic exposure, periodic selection may establish such a gene on the chromosome. The AmpC serine ß-lactamase may be a case in point and several multidrug efflux systems are found in chromosomal contexts (43). Like the serine ß-lactamases, multidrug efflux is also widespread among plasmids. Perhaps these ready made detoxification adaptations are utilized by parasitic DNAs in habitats where the levels of toxic exposure prevent the long term persistence of the organisms from which the detoxification genes were derived.

The introduction of antibiotics into clinical practice has created a novel habitat where pre-existing resistance determinants are selected strongly. Some transposons, known as

integrons, have the ability to acquire various kinds of antibiotic resistance determinants, possibly as an adaptation to the presence of multiple antibiotics in the natural environment, and integrons are now perilous agents for the dissemination of multiple antibiotic resistance in clinical environments (44). In the case of pathogenicity islands, Finlay and Falkow note (45), "Recent information suggests that pathogenic bacteria evolved from related non-pathogenic organisms by genetically acquiring relatively large blocks of genetic material rather than by slow, adaptive evolution of preexisting genes." While some determinants for invasion of eukaryotic cells probably descended from eukaryotes, a disproportionate number of genes that encode antibiotic resistance and virulence were probably derived from relatively benign antecedents on the bacterial chromosome. Perhaps we should look for potential sources of antibiotic resistance and virulence when we study the bacterial chromosome. The stable incorporation of resistance determinants into bacterial chromosomes will make it difficult to reverse antibiotic resistance in the clinic (46), but this can be avoided by prudent antibiotic usage.

References:

(1) Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life*, Murray, London.

(2) Huxley, J.S. (1942) *Evolution, the Modern Synthesis*, Allen and Unwin, London.

(3) Williams, G.C. (1966) *Adaptation and Natural Selection*, Princeton University Press, Princeton, New Jersey.

(4) Mayr, E. (1982) *The Growth of Biological Thought: diversity, evolution, and inheritance*, Harvard University Press, Cambridge, MS.

(5) Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press.

(6) Kimura, M. (1986) *Phil. Trans. R. Soc. Lond.* B **312:** 343

(7) Medawar, P.B. (1981) *New York Review of Books* **28 (2):** 34

(8) Dawkins, R. (1982) *The Extended Phenotype*, Oxford University Press, Oxford, U.K.

(9) Woese, C.R. (1994) *Microbiol. Rev.* **58:** 1

(10) Benzer, S. (1957) The Elementary Units of Heredity, *in The Chemical Basis of Heredity*, The Johns Hopkins Press, Baltimore.

(11) Jacob, F. (1973) *The Logic of Life*, Pantheon Books, New York.

(12) Stokes, H.W., and Hall, R.M. (1989) *Molec. Microbiol.* **3:** 1669

(13) Jacob, F., Brenner, S., and Cuzin F. (1963) *in Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 28, The Biological Laboratory, L.I., New York.

(14) Fristrom, J.W., and Spieth P.T. (1980) *Principles of Genetics*, Chiron Press inc., New York.

(15) Maynard-Smith, J. (1996) Population Genetics: an Introduction, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition*, ASM Press, Washington, D.C.

(16) Slatkin, M. (1972) *Genetics* **72:** 157

(17) Milkman, R. (1996) Recombinational Exchange Among Clonal Populations, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition*, ASM Press, Washington, D.C.

(18) Attwood, K.C., Schneider, L.K., and Ryan, F.J. (1951) *Proc. Natl. Acad. Sci. USA* **37:** 146

(19) Levin, B.R., and Lenski, R.E. (1983) *in Coevolution*, D.J. Futuyama and M. Slatkin eds., Sinauer Associates Inc., Sunderland, MS.

(20) Doolittle, W.F., and Sapienza, C. (1980) *Nature* **284:** 601

(21) Campbell, A. (1981) *Ann. Rev. Microbiol.* **35:** 55

(22) Hartl, D.L., Medhora, M., Green, L., and Dykhuizen, D.E. (1986) *Phil. Trans. R. Soc. Lond.* B **312:** 191

(23) Höltje, J.-V. (1995) *Arch. Microbiol.* **164:** 243

(24) Luria, S.E., and Suit, J.L. (1987) Colicins and Col Plasmids, *in Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, ASM Press, Washington, D.C.

(25) Helinski, D.R., Toukdarian, A.E., and Novick, R.P. (1996) Replication Control and Other Stable Maintenance Mechanisms of Plasmids, *in Escherichia coli and Salmonella: cellular and molecular biology, 2nd edition*, F.C. Neidhardt, ed., ASM Press, Washington, D.C.

(26) Naito, T., Kusano, K., and Kobayashi, I. (1995) *Science* **267:** 897

(27) Einhellig, F.A. (1995) *in Allelopathy: Organisms, Processes, and Applications*, Inderjit, K.M.M. Dakshini, F.A. Einhellig, eds., American Chemical Society, Washington, D.C.

(28) Demain, A.L. (1974) *Annals New York Acad. Sci.* **235:** 601

(29) Demain, A.L. (1989) Functions of Secondary Metabolites, *in Genetics and Molecular Biology of Industrial Microorganisms*, C.L. Hershberger, S.W. Queener, G. Hegeman, eds., ASM Press, Washington, D.C.

(30) Davies, J. (1990) *Molec. Microbiol.* **4:** 1227

(31) Cullum, J., Flett, F., and Piendl, W. (1989) Genetic Instability, Deletions, and DNA Amplifications in *Streptomyces* Species, *in Genetics and Molecular Biology of Industrial Microorganisms*, C.L. Hershberger, S.W. Queener, G. Hegeman, eds., ASM Press, Washington, D.C.

(32) Jensen, S.E., Leskiw, B.K., Doran, J.L., Petrich, A.K., and Westlake, D.W.S. (1989) Expression of the *Streptomyces clavuligeris* Isopenicillin N Synthase Gene in *Escherichia coli* and *Streptomyces lividans*, *in Genetics and Molecular Biology of Industrial Microorganisms*, C.L. Hershberger, S.W. Queener, G. Hegeman, eds., ASM Press, Washington, D.C.

(33) Miller, J.R., and Ingolia, T.D. (1989) Cloning ß-Lactam Genes from *Streptomyces* spp. and Fungi, *in Genetics and Molecular Biology of Industrial Microorganisms*, C.L. Hershberger, S.W. Queener, G. Hegeman, eds., ASM Press, Washington, D.C.

(34) Höltje, J.-V, and Tuomanen, E.I. (1991) *J. Gen. Microbiol.* **137:** 441

(35) Hacker, J., Blum-Oehler, G., Mühldoefer, I., and Tschäpe, H. (1997) *Molec. Microbiol.* **23:** 1089

(36) Ghuysen, J.-M (1991) *Ann. Rev. Microbiol.* **45:** 37

(37) Jaurin, B., and Grundström, T. (1981) *Proc. Natl. Acad. Sci. USA* **78:** 4897

(38) Sutcliffe, J.G. (1978) *Proc. Natl. Acad. Sci. USA* **75:** 3737

(39) Heffron, F., Sublett, R., Hedges, R.W., Jocob, A., and Falkow, S. (1975) *J. Bacteriol.* **122:** 250

(40) Watanabe, T. (1971) *Annals New York Acad. Sci.* **182:** 126

(41) Datta, N., and Hughes, V.M. (1983) *Nature* **306:** 616

(42) Levin, B.R. (1996) *Emerg. Infect. Dis.*, **2:** 93

(43) Nikaido, H. (1994) *Science* **264:** 382

(44) Davies, J. (1994) *Science*, **264:** 375

(45) Finlay, B.B., and Falkow, S. (1997) *Microbiol. and Molec. Biol. Rev.* **61:** 136

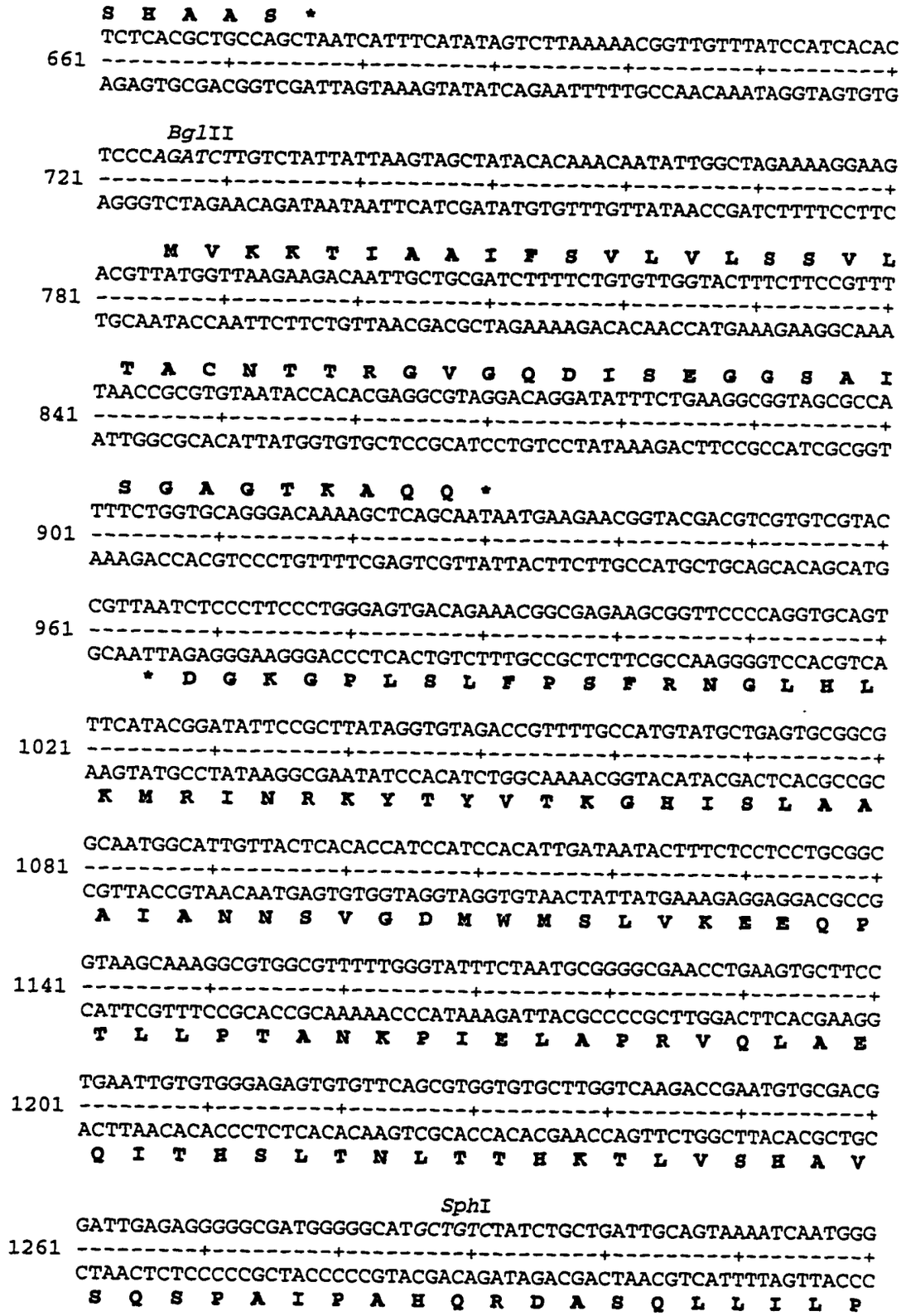
(46) Spratt, B.G. (1996) *Current Biology* **6:** 1219

APPENDIX

DNA Sequence of the *frd-efp* Gene Interval of *Citrobacter freundii*

The double stranded sequencing strategy outlined in figure 5-6 yielded DNA sequence that starts at the *Eco*RI site in the truncated *efp* operon, and extends through to the *blc* gene. This sequence will be updated in GenBank file U21727. The sequence of *ampC* and *ampR* from GenBank files X03866 and M27222 was derived from the same clone (*Citrobacter freundii* OS60) and has been incorporated into the following figure as a continuous sequence that extends to the end of *frdD*. Nucleotides are numbered to correspond approximately with the following open reading frames: *efp* (1-481); *ecnA* (541-661); *ecnB* (781-901); *ecnR* (961-1561); *sugE* (1801-2101); *blc* (2101-2641); *ampC* (2761-3901); *ampR* (4021-4861); *frdD* (4921-4980).

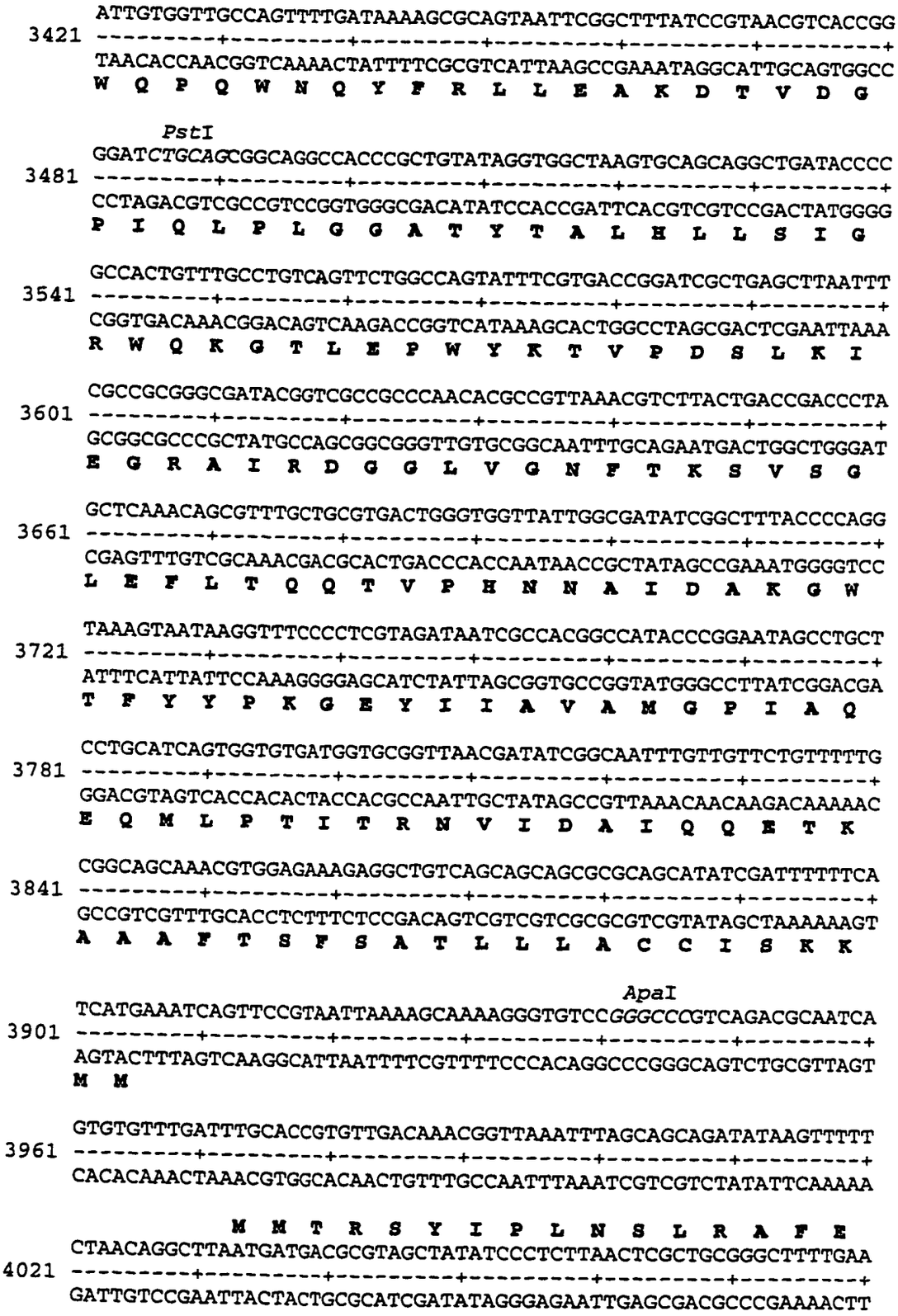
```
    E  F  V  K  P  G  K  G  Q  A  F  A  R  V  K  L  R  R  L  L
    GAATTCGTTAAACCAGGTAAAGGCCAGGCGTTTGCGCGCGTTAAGCTGCGTCGCCTGCTG
1   ---------+---------+---------+---------+---------+---------+
    CTTAAGCAATTTGGTCCATTTCCGGTCCGCAAACGCGCGCAATTCGACGCAGCGGACGAC
     EcoRI

    T  G  T  R  V  E  K  T  F  K  S  T  D  S  A  E  G  A  D  V
    ACCGGTACTCGCGTAGAGAAAACCTTCAAGTCTACTGACTCCGCTGAAGGCGCAGACGTA
61  ---------+---------+---------+---------+---------+---------+
    TGGCCATGAGCGCATCTCTTTTGGAAGTTCAGATGACTGAGGCGACTTCCGCGTCTGCAT

    A  D  M  N  L  T  F  L  Y  S  D  G  E  F  W  H  F  M  N  N
    GCCGACATGAACCTGACTTTCCTGTATAGCGACGGTGAGTTCTGGCACTTTATGAATAAC
121 ---------+---------+---------+---------+---------+---------+
    CGGCTGTACTTGGACTGAAAGGACATATCGCTGCCACTCAAGACCGTGAAATACTTATTG

    T  T  F  E  Q  L  A  A  D  E  K  A  V  G  D  N  A  K  W  L
    ACCACTTTCGAACAGCTGGCTGCTGACGAGAAAGCGGTAGGCGATAACGCTAAATGGCTG
181 ---------+---------+---------+---------+---------+---------+
    TGGTGAAAGCTTGTCGACCGACGACTGCTCTTTCGCCATCCGCTATTGCGATTTACCGAC

    L  D  Q  A  E  C  I  V  T  L  W  N  G  Q  P  I  S  V  T  P
    CTGGACCAGGCTGAATGCATCGTGACCCTGTGGAACGGTCAGCCGATCTCTGTAACTCCG
241 ---------+---------+---------+---------+---------+---------+
    GACCTGGTCCGACTTACGTAGCACTGGGACACCTTGCCAGTCGGCTAGAGACATTGAGGC

    P  N  F  V  E  L  E  V  V  E  T  D  P  G  L  K  G  D  T  A
    CCGAACTTCGTTGAGCTGGAAGTTGTTGAAACCGACCCAGGTCTGAAAGGCGATACCGCT
301 ---------+---------+---------+---------+---------+---------+
    GGCTTGAAGCAACTCGACCTTCAACAACTTTGGCTGGGTCCAGACTTTCCGCTATGGCGA

    G  T  G  G  K  P  A  T  L  S  T  G  A  V  V  K  V  P  L  F
    GGTACTGGCGGCAAACCGGCTACCCTGTCTACCGGCGCAGTCGTTAAAGTTCCGCTGTTC
361 ---------+---------+---------+---------+---------+---------+
    CCATGACCGCCGTTTGGCCGATGGGACAGATGGCCGCGTCAGCAATTTCAAGGCGACAAG

    V  Q  I  G  E  V  I  K  V  D  T  R  S  G  E  Y  V  S  R  V
    GTACAGATCGGCGAAGTGATCAAAGTGGATACCCGCTCCGGCGAATACGTTTCTCGCGTA
421 ---------+---------+---------+---------+---------+---------+
    CATGTCTAGCCGCTTCACTAGTTTCACCTATGGGCGAGGCCGCTTATGCAAAGAGCGCAT

    K  *
    AAATAAGTCTTCCATTTGATCTTATGGCGCAGGTTGCTGCGCCATACTTTTTCCAGGGCA
481 ---------+---------+---------+---------+---------+---------+
    TTTATTCAGAAGGTAAACTAGAATACCGCGTCCAACGACGCGGTATGAAAAAGGTCCCGT

      M  M  K  R  L  L  G  L  V  M  L  L  L  F  T  C  T  L  L
    GGAATGATGAAACGTCTTCTCGGTCTTGTAATGCTACTCCTTTTCACCTGCACCCTCCTG
541 ---------+---------+---------+---------+---------+---------+
    CCTTACTACTTTGCAGAAGAGCCAGAACATTACGATGAGGAAAAGTGGACGTGGGAGGAC

    T  G  C  N  T  A  R  G  F  G  E  D  I  K  H  L  G  N  S  I
    ACGGGATGCAATACCGCGCGCGGCTTCGGTGAAGACATCAAACATCTTGGCAATTCAATC
601 ---------+---------+---------+---------+---------+---------+
    TGCCCTACGTTATGGCGCGCGCCGAAGCCACTTCTGTAGTTTGTAGAACCGTTAAGTTAG
```


S H A A S *
TCTCACGCTGCCAGCTAATCATTTCATATAGTCTTAAAAACGGTTGTTTATCCATCACAC
661
AGAGTGCGACGGTCGATTAGTAAAGTATATCAGAATTTTTGCCAACAAATAGGTAGTGTG
BglII
TCCCAGATCTTGTCTATTATTAAGTAGCTATACACAAACAATATTGGCTAGAAAAGGAAG
721
AGGGTCTAGAACAGATAATAATTCATCGATATGTGTTTGTTATAACCGATCTTTTCCTTC
M V K K T I A A I F S V L V L S S V L
ACGTTATGGTTAAGAAGACAATTGCTGCGATCTTTTCTGTGTTGGTACTTTCTTCCGTTT
781
TGCAATACCAATTCTTCTGTTAACGACGCTAGAAAAGACACAACCATGAAAGAAGGCAAA
T A C N T T R G V G Q D I S E G G S A I
TAACCGCGTGTAATACCACACGAGGCGTAGGACAGGATATTTCTGAAGGCGGTAGCGCCA
841
ATTGGCGCACATTATGGTGTGCTCCGCATCCTGTCCTATAAAGACTTCCGCCATCGCGGT
S G A G T K A Q Q *
TTTCTGGTGCAGGGACAAAAGCTCAGCAATAATGAAGAACGGTACGACGTCGTGTCGTAC
901
AAAGACCACGTCCCTGTTTTCGAGTCGTTATTACTTCTTGCCATGCTGCAGCACAGCATG
CGTTAATCTCCCTTCCCTGGGAGTGACAGAAACGGCGAGAAGCGGTTCCCCAGGTGCAGT
961
GCAATTAGAGGGAAGGGACCCTCACTGTCTTTGCCGCTCTTCGCCAAGGGGTCCACGTCA
* D G K G P L S L F P S F R N G L H L
TTCATACGGATATTCCGCTTATAGGTGTAGACCGTTTTGCCATGTATGCTGAGTGCGGCG
1021
AAGTATGCCTATAAGGCGAATATCCACATCTGGCAAAACGGTACATACGACTCACGCCGC
K M R I N R K Y T Y V T K G H I S L A A
GCAATGGCATTGTTACTCACACCATCCATCCACATTGATAATACTTTCTCCTCCTGCGGC
1081
CGTTACCGTAACAATGAGTGTGGTAGGTAGGTGTAACTATTATGAAAGAGGAGGACGCCG
A I A N N S V G D M W M S L V K E E Q P
GTAAGCAAAGGCGTGGCGTTTTTGGGTATTTCTAATGCGGGGCGAACCTGAAGTGCTTCC
1141
CATTCGTTTCCGCACCGCAAAAACCCATAAAGATTACGCCCCGCTTGGACTTCACGAAGG
T L L P T A N K P I E L A P R V Q L A E
TGAATTGTGTGGGAGAGTGTGTTCAGCGTGGTGTGCTTGGTCAAGACCGAATGTGCGACG
1201
ACTTAACACACCCTCTCACACAAGTCGCACCACACGAACCAGTTCTGGCTTACACGCTGC
Q I T H S L T N L T T H K T L V S H A V
SphI
GATTGAGAGGGGGCGATGGGGGCATGCTGTCTATCTGCTGATTGCAGTAAAATCAATGGG
1261
CTAACTCTCCCCCGCTACCCCCGTACGACAGATAGACGACTAACGTCATTTTAGTTACCC
S Q S P A I P A H Q R D A S Q L L I L P

```
     ATGACGCCACAACTCATAAAAAAAGGGTTAATTGGCGGACTGGAGTGCGTCTCGCTAATA
1321 ---------+---------+---------+---------+---------+---------+
     TACTGCGGTGTTGAGTATTTTTTTCCCAATTAACCGCCTGACCTCACGCAGAGCGATTAT
      I  V  G  C  S  M  F  F  P  N  I  P  P  S  S  R  T  E  S  I

     AAACTGTTTAAATCCGCGATCACCAGATGCGGTTTCCATTGCAAAATATGCTCTTTTGCT
1381 ---------+---------+---------+---------+---------+---------+
     TTTGACAAATTTAGGCGCTAGTGGTCTACGCCAAAGGTAACGTTTTATACGAGAAAACGA
      F  S  N  L  D  A  I  V  L  H  P  K  W  Q  L  I  H  E  K  A

     AACAGGAGATTATTGATTCCTGATACCAGAAATGAAGAGCTGAACCTGCCAGAATGATTA
1441 ---------+---------+---------+---------+---------+---------+
     TTGTCCTCTAATAACTAAGGACTATGGTCTTTACTTCTCGACTTGGACGGTCTTACTAAT
      L  L  L  N  N  I  G  S  V  L  F  S  S  S  F  R  G  S  H  N

     AGCAAGGCCTCTATCCCTGTGCGGGTAAAGTGGCACCGGTCAATCACTAAAATTTTAAGC
1501 ---------+---------+---------+---------+---------+---------+
     TCGTTCCGGAGATAGGGACACGCCCATTTCACCGTGGCCAGTTAGTGATTTTAAAATTCG
      L  L  A  E  I  G  T  R  T  F  H  C  R  D  I  V  L  I  K  L

     ATAAACGTTTCGCATTCAGAAGAGGGGCAATCGACTTTCCGGTTTCAGAAAGCCCTCCAT
1561 ---------+---------+---------+---------+---------+---------+
     TATTTGCAAAGCGTAAGTCTTCTCCCCGTTAGCTGAAAGGCCAAAGTCTTTCGGGAGGTA
      M

     CATAGAGAACTCCGATGAGTGCCAAATGCGGGAACTACGCCGCTAATTATGCTTAATTCT
1621 ---------+---------+---------+---------+---------+---------+
     GTATCTCTTGAGGCTACTCACGGTTTACGCCCTTGATGCGGCGATTAATACGAATTAAGA

     TGCTTTTAGCTGGCAGTGAATAAAAAGGTTGAAATTCATTTTCTTCCGGGCAAAAATGCC
1681 ---------+---------+---------+---------+---------+---------+
     ACGAAAATCGACCGTCACTTATTTTTCCAACTTTAAGTAAAAGAAGGCCCGTTTTTACGG

     ATTTCACACATCAGCTTTGTAGGACGACCTGCAAACGCTTCTTTCACCGGGGACGGCCCC
1741 ---------+---------+---------+---------+---------+---------+
     TAAAGTGTGTAGTCGAAACATCCTGCTGGACGTTTGCGAAGAAAGTGGCCCCTGCCGGGG

                        M  S  W  I  V  L  L  I  A  G  L  L  E  V
     ATTGTTCTGGAGCCTGATATGTCCTGGATCGTTTTATTAATTGCAGGTTTGCTCGAAGTC
1801 ---------+---------+---------+---------+---------+---------+
     TAACAAGACCTCGGACTATACAGGACCTAGCAAAATAATTAACGTCCAAACGAGCTTCAG

      V  W  A  I  G  L  K  Y  T  H  G  F  T  R  L  T  P  S  I  I
     GTCTGGGCGATTGGCCTGAAATATACCCACGGTTTTACGCGTCTTACGCCAAGCATTATC
1861 ---------+---------+---------+---------+---------+---------+
     CAGACCCGCTAACCGGACTTTATATGGGTGCCAAAATGCGCAGAATGCGGTTCGTAATAG

      T  I  A  A  M  I  V  S  I  A  M  L  S  W  A  M  R  T  L  P
     ACTATTGCGGCGATGATCGTCAGTATTGCCATGCTCTCCTGGGCAATGCGCACATTACCT
1921 ---------+---------+---------+---------+---------+---------+
     TGATAACGCCGCTACTAGCAGTCATAACGGTACGAGAGGACCCGTTACGCGTGTAATGGA

      V  G  T  A  Y  A  V  W  T  G  I  G  A  V  G  A  A  I  T  G
     GTAGGAACCGCTTATGCGGTCTGGACCGGCATTGGCGCCGTCGGGGCGGCCATTACAGGG
1981 ---------+---------+---------+---------+---------+---------+
     CATCCTTGGCGAATACGCCAGACCTGGCCGTAACCGCGGCAGCCCCGCCGGTAATGTCCC
```

```
     I  L  L  L  G  E  S  A  S  P  A  R  L  L  S  L  G  L  I  V
     ATTTTGCTGCTGGGTGAATCTGCCAGCCCGGCACGTTTGCTGAGCCTTGGGCTGATCGTT
2041 ---------+---------+---------+---------+---------+---------+
     TAAAACGACGACCCACTTAGACGGTCGGGCCGTGCAAACGACTCGGAACCCGACTAGCAA

     A  G  I  I  G  L  K  L  S  T  H  *
     GCTGGCATTATTGGTCTGAAGCTGAGCACTCACTAAGCGCCAGGTTGTTTCACCCAAATC
2101 ---------+---------+---------+---------+---------+---------+
     CGACCGTAATAACCAGACTTCGACTCGTGAGTGATTCGCGGTCCAACAAAGTGGGTTTAG
                                        *  A  G  P  Q  K  V  W  I

                                   SmaI
     AGCTTATTCACCTCAAACCCTTCCCGGGTCGCGATGGCTAACATTTGCTGTTTCATTTCA
2161 ---------+---------+---------+---------+---------+---------+
     TCGAATAAGTGGAGTTTGGGAAGGGCCCAGCGCTACCGATTGTAAACGACAAAGTAAAGT
      L  K  N  V  E  F  G  E  R  T  A  I  A  L  M  Q  Q  K  M  E

     TCTGAAATAGTTGGGGTCCGTGAAAGGATCCACAGGTAGTCGCGATCCGGACCGCAAACC
2221 ---------+---------+---------+---------+---------+---------+
     AGACTTTATCAACCCCAGGCACTTTCCTAGGTGTCCATCAGCGCTAGGCCTGGCGTTTGG
      D  S  I  T  P  T  R  S  L  I  W  L  Y  D  R  D  P  G  C  V

     AGCGCGTGACGATATTCCCGGTCGAGTGCAATTACGTTATACCCACCATAGAAGGGGCCG
2281 ---------+---------+---------+---------+---------+---------+
     TCGCGCACTGCTATAAGGGCCAGCTCACGTTAATGCAATATGGGTGGTATCTTCCCCGGC
      L  A  H  R  Y  E  R  D  L  A  I  V  N  Y  G  G  Y  F  P  G

     AAGAATGATACTTTAAGCGCGGCCGTGCTTGGGTCGCCGGTGAAATAGGCTTTCCCTTCC
2341 ---------+---------+---------+---------+---------+---------+
     TTCTTACTATGAAATTCGCGCCGGCACGAACCCAGCGGCCACTTTATCCGAAAGGGAAGG
      F  F  S  V  K  L  A  A  T  S  P  D  G  T  F  Y  A  K  G  E

     GTTTTCTGCCACATCTCCCTGTCAGGGTTATAGCCCTTGTTAATAACGTTGATGCCGCCG
2401 ---------+---------+---------+---------+---------+---------+
     CAAAAGACGGTGTAGAGGGACAGTCCCAATATCGGGAACAATTATTGCAACTACGGCGGC
      T  K  Q  W  M  E  R  D  P  N  Y  G  K  N  I  V  N  I  G  G

     TCGTCGCGCAAGCTGTACGTTGCGGTCACTTTATCCAACCCGCGCTCGAAACGATGGTCG
2461 ---------+---------+---------+---------+---------+---------+
     AGCAGCGCGTTCGACATGCAACGCCAGTGAAATAGGTTGGGCGCGAGCTTTGCTACCAGC
      D  D  R  L  S  Y  T  A  T  V  K  D  L  G  R  E  F  R  H  D

     AAGCGCGCAATTTCATACCATGTTCCCAGATAGCGTTTGGCATCAAAGTTATTCACCACG
2521 ---------+---------+---------+---------+---------+---------+
     TTCGCGCGTTAAAGTATGGTACAAGGGTCTATCGCAAACCGTAGTTTCAATAAGTGGTGC
      F  R  A  I  E  Y  W  T  G  L  Y  R  K  A  D  F  N  N  V  V

     GTAACGCCTTTCGGCGGTGTTGGGGAGCTGCAAGCGACAACCAGGAATGCAGCCGTAACT
2581 ---------+---------+---------+---------+---------+---------+
     CATTGCGGAAAGCCGCCACAACCCCTCGACGTTCGCTGTTGGTCCTTACGTCGGCATTGA
      T  V  G  K  P  P  T  P  S  S  C  A  V  V  L  F  A  A  T  V

      PstI
     GCAGCAACGACGGGCAGAATGCGCATAGTGTTTTCCTTACTGTTTTTTGTTAAGTGTAGA
2641 ---------+---------+---------+---------+---------+---------+
     CGTCGTTGCTGCCCGTCTTACGCGTATCACAAAAGGAATGACAAAAAACAATTCACATCT
      A  A  V  V  P  L  I  R  M
```

```
     TGACAGCAGGGAAAAAAGAGCAAGAAAGGAGGTCCGATTTACCGGACCCCTCTGTCAGTT
2701 ---------+---------+---------+---------+---------+---------+
     ACTGTCGTCCCTTTTTTCTCGTTCTTTCCTCCAGGCTAAATGGCCTGGGGAGACAGTCAA

     ATTGCAGTTTTTCAAGAATGCGCCAGGCCGCCTCGACGCGAGCCGGGTTGGGGTAGCTTT
2761 ---------+---------+---------+---------+---------+---------+
     TAACGTCAAAAAGTTCTTACGCGGTCCGGCGGAGCTGCGCTCGGCCCAACCCCATCGAAA
     *  Q  L  K  E  L  I  R  W  A  A  E  V  R  A  P  N  P  Y  S

     TGTTTGCCAACATTACGATGCCAAGGTTTTTTTCTGGAACGAAAGCAACGTAGCTGCCAA
2821 ---------+---------+---------+---------+---------+---------+
     ACAAACGGTTGTAATGCTACGGTTCCAAAAAAAGACCTTGCTTTCGTTGCATCGACGGTT
     K  N  A  L  M  V  I  G  L  N  K  E  P  V  F  A  V  Y  S  G

     ATCCGCCTGTGGATCCTGTTTTATGCACCCATGAGGCTTTCACGGCAGGTGCTGGCGGGT
2881 ---------+---------+---------+---------+---------+---------+
     TAGGCGGACACCTAGGACAAAATACGTGGGTACTCCGAAAGTGCCGTCCACGACCGCCCA
     F  G  G  T  S  G  T  K  H  V  W  S  A  K  V  A  P  A  P  P

     TTACCTCAACGGCGGGAAGCGCTGCCAATGCCACTTTGCTGTCGCTGCCGTTGATGATCG
2941 ---------+---------+---------+---------+---------+---------+
     AATGGAGTTGCCGCCCTTCGCGACGGTTACGGTGAAACGACAGCGACGGCAACTACTAGC
     N  V  E  V  A  P  L  A  A  L  A  V  K  S  D  S  G  N  I  I

     AATCAGCTTTCAGCGGCCAGTTCAGCATCTCCCAGCCTAATCCCTGGTACATATCACCAA
3001 ---------+---------+---------+---------+---------+---------+
     TTAGTCGAAAGTCGCCGGTCAAGTCGTAGAGGGTCGGATTAGGGACCATGTATAGTGGTT
     S  D  A  K  L  P  W  N  L  M  E  W  G  L  G  Q  Y  M  D  G

     TACGCCAGTAGCGAGACTGCGCAAGCTCAATGCCCTGCTGGAGCGTTTTCTCCTGAACGT
3061 ---------+---------+---------+---------+---------+---------+
     ATGCGGTCATCGCTCTGACGCGTTCGAGTTACGGGACGACCTCGCAAAAGAGGACTTGCA
     I  R  W  Y  R  S  Q  A  L  E  I  G  Q  Q  L  T  K  E  Q  V

     GGCTGGCGTCCATGTTGGCCTGAACCCAGCGGGCCATATCGATAACGCTGGATTTCACGC
3121 ---------+---------+---------+---------+---------+---------+
     CCGACCGCAGGTACAACCGGACTTGGGTCGCCCGGTATAGCTATTGCGACCTAAAGTGCG
     H  S  A  D  M  N  A  Q  V  W  R  A  M  D  I  V  S  S  K  V

     CATAGGCTTCGGCGTCAAGTTGTCCCGGAGAAACGTGCACAGGCTTCCCTTCGAGATAGC
3181 ---------+---------+---------+---------+---------+---------+
     GTATCCGAAGCCGCAGTTCAACAGGGCCTCTTTGCACGTGTCCGAAGGGAAGCTCTATCG
     G  Y  A  E  A  D  L  Q  G  P  S  V  H  V  P  K  G  E  L  Y

     CCCAGGCATAGTTTTTTTGTTCGCTTTGCGGAACCGTAATCCAGGTATGCGCCAGTTTTA
3241 ---------+---------+---------+---------+---------+---------+
     GGGTCCGTATCAAAAAAACAAGCGAAACGCCTTGGCATTAGGTCCATACGCGGTCAAAAT
     G  W  A  Y  N  K  Q  E  S  Q  P  V  T  I  W  T  H  A  L  K

     ATGGTTGCAGGACGCGTCTGGTCATTGCCTCTTCGTAGCTCATACCTGAAGATTTCACCG
3301 ---------+---------+---------+---------+---------+---------+
     TACCAACGTCCTGCGCAGACCAGTAACGGAGAAGCATCGAGTATGGACTTCTAAAGTGGC
     L  P  Q  L  V  R  R  T  M  A  E  E  Y  S  M  G  S  S  K  V

     CCAGCGCACCAAACAGACCAATGCTGGAGTTAGCGTAAAGACGCTTAGCGCCCGGAGTCC
3361 ---------+---------+---------+---------+---------+---------+
     GGTCGCGTGGTTTGTCTGGTTACGACCTCAATCGCATTTCTGCGAATCGCGGGCCTCAGG
     A  L  A  G  F  L  G  I  S  S  N  A  Y  L  R  K  A  G  P  T
```


ATTGTGGTTGCCAGTTTTGATAAAAGCGCAGTAATTCGGCTTTATCCGTAACGTCACCGG
3421
TAACACCAACGGTCAAAACTATTTTCGCGTCATTAAGCCGAAATAGGCATTGCAGTGGCC
W Q P Q W N Q Y F R L L E A K D T V D G
PstI
GGATCTGCAGCGGCAGGCCACCCGCTGTATAGGTGGCTAAGTGCAGCAGGCTGATACCCC
3481
CCTAGACGTCGCCGTCCGGTGGGCGACATATCCACCGATTCACGTCGTCCGACTATGGGG
P I Q L P L G G A T Y T A L H L L S I G
GCCACTGTTTGCCTGTCAGTTCTGGCCAGTATTTCGTGACCGGATCGCTGAGCTTAATTT
3541
CGGTGACAAACGGACAGTCAAGACCGGTCATAAAGCACTGGCCTAGCGACTCGAATTAAA
R W Q K G T L E P W Y K T V P D S L K I
CGCCGCGGGCGATACGGTCGCCGCCCAACACGCCGTTAAACGTCTTACTGACCGACCCTA
3601
GCGGCGCCCGCTATGCCAGCGGCGGGTTGTGCGGCAATTTGCAGAATGACTGGCTGGGAT
E G R A I R D G G L V G N F T K S V S G
GCTCAAACAGCGTTTGCTGCGTGACTGGGTGGTTATTGGCGATATCGGCTTTACCCCAGG
3661
CGAGTTTGTCGCAAACGACGCACTGACCCACCAATAACCGCTATAGCCGAAATGGGGTCC
L E F L T Q Q T V P H N N A I D A K G W
TAAAGTAATAAGGTTTCCCCTCGTAGATAATCGCCACGGCCATACCCGGAATAGCCTGCT
3721
ATTTCATTATTCCAAAGGGGAGCATCTATTAGCGGTGCCGGTATGGGCCTTATCGGACGA
T F Y Y P K G E Y I I A V A M G P I A Q
CCTGCATCAGTGGTGTGATGGTGCGGTTAACGATATCGGCAATTTGTTGTTCTGTTTTTG
3781
GGACGTAGTCACCACACTACCACGCCAATTGCTATAGCCGTTAAACAACAAGACAAAAAC
E Q M L P T I T R N V I D A I Q Q E T K
CGGCAGCAAACGTGGAGAAAGAGGCTGTCAGCAGCAGCGCGCAGCATATCGATTTTTTCA
3841
GCCGTCGTTTGCACCTCTTTCTCCGACAGTCGTCGTCGCGCGTCGTATAGCTAAAAAAGT
A A A F T S F S A T L L L A C C I S K K
ApaI
TCATGAAATCAGTTCCGTAATTAAAAGCAAAAGGGTGTCCGGGCCCGTCAGACGCAATCA
3901
AGTACTTTAGTCAAGGCATTAATTTTCGTTTTCCCACAGGCCCGGGCAGTCTGCGTTAGT
M M
GTGTGTTTGATTTGCACCGTGTTGACAAACGGTTAAATTTAGCAGCAGATATAAGTTTTT
3961
CACACAAACTAAACGTGGCACAACTGTTTGCCAATTTAAATCGTCGTCTATATTCAAAAA
M M T R S Y I P L N S L R A F E
CTAACAGGCTTAATGATGACGCGTAGCTATATCCCTCTTAACTCGCTGCGGGCTTTTGAA
4021
GATTGTCCGAATTACTACTGCGCATCGATATAGGGAGAATTGAGCGACGCCCGAAAACTT

```
     A  A  A  R  H  L  S  F  T  R  A  A  I  E  L  N  V  T  H  S
     GCCGCGGCCAGACATCTCAGCTTTACCCGCGCTGCGATTGAGCTCAACGTGACGCATTCT
4081 ---------+---------+---------+---------+---------+---------+
     CGGCGCCGGTCTGTAGAGTCGAAATGGGCGCGACGCTAACTCGAGTTGCACTGCGTAAGA

     A  I  S  Q  H  V  K  S  L  E  Q  Q  L  N  C  Q  L  F  V  R
     GCCATCAGCCAGCATGTCAAATCGCTGGAGCAGCAGCTGAATTGTCAGCTTTTTGTGCGC
4141 ---------+---------+---------+---------+---------+---------+
     CGGTAGTCGGTCGTACAGTTTAGCGACCTCGTCGTCGACTTAACAGTCGAAAAACACGCG

     G  S  R  G  L  M  L  T  T  E  G  E  S  L  L  P  V  L  N  D
     GGTTCTCGTGGATTAATGTTGACGACTGAAGGGGAAAGTCTGTTACCTGTACTTAATGAC
4201 ---------+---------+---------+---------+---------+---------+
     CCAAGAGCACCTAATTACAACTGCTGACTTCCCCTTTCAGACAATGGACATGAATTACTG

     S  F  D  R  M  A  G  M  L  D  R  F  A  T  K  Q  T  Q  E  K
     TCCTTCGATCGTATGGCGGGGATGTTGGATCGTTTTGCCACTAAACAGACCCAGGAAAAA
4261 ---------+---------+---------+---------+---------+---------+
     AGGAAGCTAGCATACCGCCCCTACAACCTAGCAAAACGGTGATTTGTCTGGGTCCTTTTT

     L  K  I  G  V  V  G  T  F  A  I  G  C  L  F  P  L  L  S  D
     CTTAAAATTGGCGTAGTGGGAACCTTTGCTATCGGCTGTCTTTTCCCGCTGTTGAGCGAT
4321 ---------+---------+---------+---------+---------+---------+
     GAATTTTAACCGCATCACCCTTGGAAACGATAGCCGACAGAAAAGGGCGACAACTCGCTA

     F  K  R  S  Y  P  H  I  D  L  H  I  S  T  H  N  N  R  V  D
     TTTAAGCGCAGTTACCCACATATTGATTTGCATATTTCTACCCATAACAATCGCGTGGAT
4381 ---------+---------+---------+---------+---------+---------+
     AAATTCGCGTCAATGGGTGTATAACTAAACGTATAAAGATGGGTATTGTTAGCGCACCTA

     P  A  A  E  G  L  D  Y  T  I  R  Y  G  G  G  A  W  H  D  T
     CCCGCCGCCGAAGGGCTGGATTATACCATTCGCTACGGTGGGGGAGCCTGGCACGATACC
4441 ---------+---------+---------+---------+---------+---------+
     GGGCGGCGGCTTCCCGACCTAATATGGTAAGCGATGCCACCCCCTCGGACCGTGCTATGG

     D  A  Q  Y  L  C  S  A  L  M  S  P  L  C  S  P  T  L  A  S
     GATGCGCAATATTTATGTAGCGCGCTGATGTCTCCACTGTGTTCGCCAACATTGGCTTCG
4501 ---------+---------+---------+---------+---------+---------+
     CTACGCGTTATAAATACATCGCGCGACTACAGAGGTGACACAAGCGGTTGTAACCGAAGC

     Q  I  Q  T  P  A  D  I  L  K  F  P  L  L  R  S  Y  R  R  D
     CAGATTCAAACGCCTGCGGATATCCTGAAATTTCCGTTATTACGATCGTATCGGCGGGAT
4561 ---------+---------+---------+---------+---------+---------+
     GTCTAAGTTTGCGGACGCCTATAGGACTTTAAAGGCAATAATGCTAGCATAGCCGCCCTA

     E  W  A  L  W  M  Q  A  A  G  E  A  P  P  S  P  T  H  N  V
     GAATGGGCGCTTTGGATGCAGGCGGCCGGAGAGGCGCCTCCGTCACCGACGCACAATGTG
4621 ---------+---------+---------+---------+---------+---------+
     CTTACCCGCGAAACCTACGTCCGCCGGCCTCTCCGCGGAGGCAGTGGCTGCGTGTTACAC

     M  V  F  D  S  S  V  T  M  L  E  A  A  Q  G  G  M  G  V  A
     ATGGTGTTTGATTCGTCAGTCACCATGTTGGAAGCTGCACAGGGGGGAATGGGTGTGGCA
4681 ---------+---------+---------+---------+---------+---------+
     TACCACAAACTAAGCAGTCAGTGGTACAACCTTCGACGTGTCCCCCCTTACCCACACCGT
```

```
     I  A  P  V  R  M  F  T  H  L  L  S  S  E  R  I  V  Q  P  F
     ATTGCGCCAGTCAGAATGTTCACGCATTTACTCAGCAGTGAACGTATCGTTCAGCCGTTT
4741 ---------+---------+---------+---------+---------+---------+
     TAACGCGGTCAGTCTTACAAGTGCGTAAATGAGTCGTCACTTGCATAGCAAGTCGGCAAA

     L  T  Q  I  D  L  G  S  Y  W  I  T  R  L  Q  S  R  P  E  T
     TTAACGCAGATTGATTTGGGAAGCTACTGGATAACGCGTTTGCAATCTCGCCCGGAGACG
4801 ---------+---------+---------+---------+---------+---------+
     AATTGCGTCTAACTAAACCCTTCGATGACCTATTGCGCAAACGTTAGAGCGGGCCTCTGC

     P  A  M  R  E  F  S  R  W  L  T  G  V  L  H  K  *
     CCCGCGATGCGTGAATTTTCCCGTTGGTTGACCGGGGTGCTGCACAAATAAAAAAGGCCC
4861 ---------+---------+---------+---------+---------+---------+
     GGGCGCTACGCACTTAAAAGGGCAACCAACTGGCCCCACGACGTGTTTATTTTTTCCGGG

     GCCAATAGCGGGCCTGATAATTAACGCATGTACTTAGATGGTCAGAATACCAATCGCCGT
4921 ---------+---------+---------+---------+---------+---------+
     CGGTTATCGCCCGGACTATTAATTGCGTACATGAATCTACCAGTCTTATGGTTAGCGGCA
                                      *  I  T  L  I  G  I  A  T
```